

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 25, 2002, 14:17:36 : Search time 517.286 Seconds
(without alignments)
54.868 Million cell updates/sec

Title: US-10-068-080-1
Perfect score: 1071
Sequence: 1 MITVLLILAVLLGSLPSGL.....NKTNLPVWGLNLTHQDPK 213

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DR seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_101002.*
1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	DB ID	Description
1	1071	100.0	213	21	AA192246 S. pneumoniae S-yn
2	1071	100.0	213	22	AAU37736 Streptococcus pneu
3	1071	100.0	213	22	AAU37943 Streptococcus pneu
4	716.5	66.9	213	23	ABB54294 Lactococcus lactis
5	716	66.9	240	23	ABP26819 Streptococcus poly
6	708	66.1	212	23	ABP26818 Streptococcus poly
7	700	65.4	207	23	ABP30349 Streptococcus poly
8	548	51.2	114	20	AA192561 Bacterial general
9	468	43.7	193	20	AA192580 B. subtilis B-ynes
10	468	43.7	193	21	AA192247 B. subtilis B-ynes

11	463.5	43.3	198	23	ABB48775
12	463.5	43.3	204	23	ABP40394
13	450.5	42.1	197	22	AAG82881
14	430.5	40.9	202	22	AAU33809
15	430.5	40.9	202	22	AAU36544
16	266.5	24.9	203	22	AAU38164
17	263.5	24.6	205	22	AAU34710
18	252.5	23.6	205	22	AAU36120
19	249.5	23.3	173	21	AAU74919
20	244.5	22.8	199	22	AAU35404
21	240	22.4	173	21	AAU74918
22	239.5	22.4	173	21	AAU74917
23	239	22.3	189	22	AAU36206
24	213	19.9	220	22	AAU36021
25	211	19.7	262	19	AAU98757
26	211	19.7	262	19	AAU35860
27	130	12.1	336	23	ABP26867
28	129	12.0	336	20	AAU14584
29	110.5	10.3	461	23	ABP26853
30	104.5	9.8	419	23	ABP28128
31	103	9.6	542	22	ABB56844
32	102.5	9.6	641	23	ABP30392
33	102.5	9.6	654	23	ABP28438
34	102	9.5	650	22	AAU37748
35	101	9.4	342	22	AAG82880
36	101	9.4	347	23	ABP39115
37	100.5	9.4	456	22	AG87158
38	100	9.3	629	22	ABB40427
39	99	9.2	629	22	AAE00297
40	98.5	9.2	388	23	ABP47596
41	98	9.2	391	23	ABP49543
42	97.5	9.1	523	23	AAE14405
43	97	9.1	179	23	ABP25497
44	97	9.1	332	22	AAH87793
45	96.5	9.0	377	23	ABP26528

ALIGNMENTS

RESULT 1
AA192246
ID AA192246 standard; Protein: 213 AA.
XX
AC AA192246;
XX
DT 10-AUG-2000 (first entry)
XX
DE S. pneumoniae S-ynes polypeptide.
XX
KW S-ynes; survival; antibacterial; inhibitor.
XX
OS Streptococcus pneumoniae.
XX
PN WO200020627-A1.
XX
PD 13-APR-2000.
XX
PF 30-SEP-1999; 99WO-US22665.
XX
PR 30-SEP-1998; 98US-0163445.
XX
PA (MILL-) MILLENNIUM PHARM INC.
XX
PI Fritz C, Youngman P, Guzman L;
DR WPI; 2000-303799/26.
DR N-PSDB; AA09181.
XX
PT Methods for identifying an antibacterial agent for treating
PT Streptococcus pneumoniae infections comprises detecting an interaction
PT between a ynes polypeptide and a test compound
XX

Listeria monocytog
Staphylococcus epi
S. epidermidis ope
Staphylococcus aur
Staphylococcus aur
Salmonella typhi c
E. coli cellular p
Klebsiella pneumon
Neisseria meningit
Haemophilus influe
Neisseria meningit
Neisseria gonorrhe
Pseudomonas aerugi
Helicobacter pylor
H. pylori GPHO 108
Helicobacter pylor
Streptococcus poly
Streptococcus poly
Group B streptococ
Streptococcus poly
Streptococcus pneu
S. epidermidis ope
Staphylococcus epi
Amino acid sequenc
Drosophila melanog
Drosophila melanog
Listeria monocytog
Listeria monocytog
Human neurotransmi
Streptococcus poly
Rat T2R13 amino ac
Streptococcus poly

PS Disclosure; Fig 1; 65pp; English.

XX Streptococcus pneumoniae yneS gene encodes a polypeptide (S-yneS)
CC essential for survival for a wide range of bacteria. Identifying an
CC antibacterial agent comprises contacting a yneS polypeptide (S-yneS) with
CC a test compound and detecting an interaction of the test compound with
CC the S-yneS polypeptide which indicates that the compound is an
CC antibacterial agent. Alternatively, detecting a decrease in function of
CC the polypeptide contacted with the test compound and determining whether
CC the compound inhibits growth of bacteria, relative to the growth of
CC bacteria cultured in the absence of a test compound where inhibition of
CC growth indicates the compound is an antibacterial agent. Inhibitors of
CC S-yneS function are useful for treating a Streptococcus pneumoniae
CC infection in mammals.

XX Sequence 213 AA;

Query Match 100.0%; Score 1071; DB 21; Length 213;
Best Local Similarity 100.0%; Pred. No. 1.1e-116; Indels 0; Gaps 0;
Matches 213; Conservative 0; Mismatches 0;

OY 1 MITIVLLILAYLLGSLGSLWIGVFFQINLREHSGNGTGTNTFRILGKKGAGMATFVID 60
DB 1 MITIVLLILAYLLGSLGSLWIGVFFQINLREHSGNGTGTNTFRILGKKGAGMATFVID 60
OY 61 FFKGTATLLPIIFHLQGVSPILFGLLAVIGHTFTPIFAGKGGKAVATSAGVIFGFAPIF 120
DB 61 FFKGTATLLPIIFHLQGVSPILFGLLAVIGHTFTPIFAGKGGKAVATSAGVIFGFAPIF 120
OY 121 CLYLAIFFGALYLGSMISLSSVTASIAAIGVLLPFLPGFILSNYDSLFIALLALASL 180
DB 121 CLYLAIFFGALYLGSMISLSSVTASIAAIGVLLPFLPGFILSNYDSLFIALLALASL 180
OY 181 IIRHKDNIARIKNTENLVPWGLNLTHODPKK 213
DB 181 IIRHKDNIARIKNTENLVPWGLNLTHODPKK 213

RESULT 2
AAU37736

ID AAU37736 standard; Protein; 213 AA.

XX AAU37736;

XX 14-FEB-2002 (first entry)

XX Streptococcus pneumoniae cellular proliferation protein #165.

DE Antisense; prokaryotic cellular proliferation protein;

KW antibiotic; antibacterial; drug design.

XX Streptococcus pneumoniae.

OS WO200170955-A2.

XX 27-SEP-2001.

XX 21-MAR-2001; 2001WO-US09180.

XX 21-MAR-2000; 2000US-191078P.

PR 23-MAY-2000; 2000US-206848P.

PR 26-MAY-2000; 2000US-20727P.

PR 23-OCT-2000; 2000US-242578P.

PR 27-NOV-2000; 2000US-253625P.

PR 22-DEC-2000; 2000US-257931P.

PR 16-FEB-2001; 2001US-269308P.

XX (ELIT-) ELITRA PHARM INC.

PA Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

XX Yamamoto RT, Xu HH;

XX WPI; 2001-611495/70.

DR

DR N-PSDB; AAS555595.

XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX Example 3; Seq ID No 13329; 511pp; English.

XX The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.

CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 213 AA;

Query Match 100.0%; Score 1071; DB 22; Length 213;

Best Local Similarity 100.0%; Pred. No. 1.1e-116; Indels 0; Gaps 0;

Matches 213; Conservative 0; Mismatches 0;

OY 1 MITIVLLILAYLLGSLGSLWIGVFFQINLREHSGNGTGTNTFRILGKKGAGMATFVID 60
DB 1 MITIVLLILAYLLGSLGSLWIGVFFQINLREHSGNGTGTNTFRILGKKGAGMATFVID 60
OY 61 FFKGTATLLPIIFHLQGVSPILFGLLAVIGHTFTPIFAGKGGKAVATSAGVIFGFAPIF 120
DB 61 FFKGTATLLPIIFHLQGVSPILFGLLAVIGHTFTPIFAGKGGKAVATSAGVIFGFAPIF 120
OY 121 CLYLAIFFGALYLGSMISLSSVTASIAAIGVLLPFLPGFILSNYDSLFIALLALASL 180
DB 121 CLYLAIFFGALYLGSMISLSSVTASIAAIGVLLPFLPGFILSNYDSLFIALLALASL 180
OY 181 IIRHKDNIARIKNTENLVPWGLNLTHODPKK 213
DB 181 IIRHKDNIARIKNTENLVPWGLNLTHODPKK 213

RESULT 3

AAU37943

ID AAU37943 standard; Protein; 213 AA.

XX AAU37943;

XX 14-FEB-2002 (first entry)

XX Streptococcus pneumoniae cellular proliferation protein #372.

DE Antisense; prokaryotic cellular proliferation protein;

KW antibiotic; antibacterial; drug design.

XX Streptococcus pneumoniae.

OS WO200170955-A2.

XX 27-SEP-2001.

XX 21-MAR-2001; 2001WO-US09180.

PR 21-MAR-2000; 2000US-191078P.

PR 23-MAY-2000; 2000US-206848P.

26-MAY-2000; 2000US-207727P.
 23-OCT-2000; 2000US-242578P.
 27-NOV-2000; 2000US-253625P.
 22-DEC-2000; 2000US-257931P.
 16-FEB-2001; 2001US-269308P.
 (ELIT-) ELITRA PHARM INC.
 Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
 Yamamoto RT, Xu HH;
 WPI: 2001-611495/70.
 N-PSDB; AAS55802.
 New polynucleotides for the identification and development of
 antibiotics, comprise sequences of antisense nucleic acids -
 Example 3; Seq ID No 13536; 511pp; English.
 The invention relates to antisense inhibitors of genes essential to
 prokaryotic cellular proliferation, their use in identifying the
 genes, their use in the discovery of novel antibiotics, the essential
 genes themselves and the encoded proteins. The prokaryotes used are
 Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
 pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
 invention is also useful for the identification of potential new targets
 for antibiotic development. The antisense nucleic acids can also be used
 to identify proteins used in proliferation, to express these proteins,
 and to obtain antibodies capable of binding to the expressed proteins.
 The proteins can be used to screen compounds in rational drug discovery
 programmes. The antisense nucleic acid sequence is also useful to screen
 for homologous nucleic acids which are required for cell proliferation in
 a wide variety of organisms. The present sequence represents an
 essential prokaryotic cellular proliferation protein.
 Note: The sequence data for this patent did not form part
 of the printed specification, but was obtained in electronic
 format directly from WIPO at
 ftp.wipo.int/pub/published_pct_sequences.

Seq Sequence 213 AA;

Query Match 100.0%; Score 1071; DB 22; Length 213;
 Best Local Similarity 100.0%; Pred. No. 1.1e-116;
 Matches 213; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MTIVLLIAYLLGSLPSGLWIGQVFFQINLREHSGNGTGTNTFFILCKKAGMATFVID 60
 DB 1 MTIVLLIAYLLGSLPSGLWIGQVFFQINLREHSGNGTGTNTFFILCKKAGMATFVID 60
 QY 61 FFKGTATLPIIFHLQGVSPILFGLLAVIGHTFPFAGFKGKAVATSGAVIFGPAPIF 120
 DB 61 FFKGTATLPIIFHLQGVSPILFGLLAVIGHTFPFAGFKGKAVATSGAVIFGPAPIF 120
 QY 121 CLYLAIFFGALYGLSMISLSSVTASIAAVIGVLLFPFLFGFILSNYDSLFATIALASL 180
 DB 121 CLYLAIFFGALYGLSMISLSSVTASIAAVIGVLLFPFLFGFILSNYDSLFATIALASL 180
 QY 181 IIRHKDNRIKKNKTENLVPWGLNTHQDPKK 213
 DB 181 IIRHKDNRIKKNKTENLVPWGLNTHQDPKK 213

RESULT 4
 ABB54294
 ID ABB54294 standard; Protein; 213 AA.
 AC ABB54294;
 XX
 XX
 DT 16-MAY-2002 (first entry)
 DE Lactococcus lactis protein ykac.
 KW Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.

XX Lactococcus lactis IL1403.
 OS
 XX FR2807446-A1.
 PN
 XX 12-OCT-2001.
 PD
 XX 11-APR-2000; 2000FR-0004630.
 PF
 XX 11-APR-2000; 2000FR-0004630.
 PR
 XX (INRG) INRA INST NAT RECH AGRONOMIQUE.
 PA
 XX Bolotine A, Sorokine A, Renault P, Ehrlich SD;
 PI
 XX WPI: 2002-043418/06.
 DR
 XX New nucleotide sequence useful in the identification of Lactococcus
 PT lactis and related species -
 PT
 XX Claim 6; SEQ ID No 996; 2504pp; French.
 PS
 XX The present invention is related to a Lactococcus lactis nucleotide
 CC sequence (ABA90521) and related proteins (ABB53300-ABB55621). The
 CC nucleic acid sequence is useful in the detection and/or amplification of
 CC nucleic acid sequence, particularly to identify Lactococcus lactis or
 CC related species. The proteins of the invention are useful for the
 CC biosynthesis or biodegradation of a composition of interest. The
 CC invention helps research in lactic bacteria, particularly useful in the
 CC production of yogurt and cheese.
 CC Note: The sequence data for this patent is based on equivalent patent
 CC WO200177334 (published 18-OCT-2001) which is available in electronic
 CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 213 AA;
 Query Match 66.9%; Score 716.5; DB 23; Length 213;
 Best Local Similarity 64.8%; Pred. No. 2.7e-75;
 Matches 138; Conservative 35; Mismatches 39; Indels 1; Gaps 1;
 QY 1 MTIVLLIAYLLGSLPSGLWIGQVFFQINLREHSGNGTGTNTFFILCKKAGMATFVI 59
 DB 1 MTIVLLIAYLLGSLPSGLWIGQVFFQINLREHSGNGTGTNTFFILCKKAGMATFVI 60
 QY 60 DFFKGTATLPIIFHLQGVSPILFGLLAVIGHTFPFAGFKGKAVATSGAVIFGPAPIF 119
 DB 61 DFLKGTATLPIIFHLQGVSPILFGLLAVIGHTFPFAGFKGKAVATSGAVIFGPAPIF 120
 QY 120 FCYLAIFFGALYGLSMISLSSVTASIAAVIGVLLFPFLFGFILSNYDSLFATIALAS 179
 DB 121 FLJYLLVPEIIVLWLFMSLSLSSVIGAVFALLGILFIPSGILFILTYSLLFIIFVLAI 180
 QY 180 LIIIRHKDNRIKKNKTENLVPWGLNTHQDPK 212
 DB 181 IIRHRRTNLKRIKHNCESLVPFGLNLSQKEK 213
 RESULT 5
 ABP26819
 ID ABP26819 standard; Protein; 240 AA.
 XX
 AC ABP26819;
 XX
 XX 02-JUL-2002 (first entry)
 DT Streptococcus polypeptide SEQ ID NO 2814.
 DE Streptococcus
 XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
 KW group A streptococcus; Streptococcus pyogenes; antibacterial;
 XX antiinflammatory; infection; vaccine; meningitis; gene therapy.
 OS Streptococcus pyogenes.
 XX

```
PN WO200234771-A2.
XX
XX
PD 02-MAY-2002.
XX
XX 29-OCT-2001; 2001WO-GB04789.
XX
XX 27-OCT-2000; 2000GB-0026333.
PR 24-NOV-2000; 2000GB-0028727.
PR 07-MAR-2001; 2001GB-0005640.
XX
XX (CHIR-) CHIRON SPA.
PA (GENO-) INST GENOMIC RES.
XX
XX Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;
PI Tettelin H;
XX
XX WPI: 2002-352536/38.
DR N-PSDB; ABN67450.
XX
XX New Streptococcus protein for the treatment or prevention of infection
PT or disease caused by Streptococcus bacteria, such as meningitis, and
PT for detecting a compound that binds to the protein -
XX
XX Claim 1; Page 3432; 4525pp; English.
XX
XX The invention relates to a protein (ABP25413-ABP30895) from group B
CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC the specification. The proteins have antibacterial and antiinflammatory
CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC antibodies that bind (I) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC Nucleic acids encoding (I) are used to detect Streptococcus in a
CC biological sample. (I) is used to determine whether a compound binds to
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC used as a vaccine or diagnostic composition. The disease caused by
CC Streptococcus that is prevented or treated may be meningitis. Nucleic
CC acid encoding (I) may be used to recombinantly produce (I) and may be
CC used in gene therapy. Antibodies to (I) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
CC Streptococcus proteins.
XX
XX Sequence 240 AA;
XX
XX Query Match 66.9%; Score 716; DB 23; Length 240;
XX Best Local Similarity 64.8%; Pred. No. 3.6e-75;
XX Matches 138; Conservative 27; Mismatches 48; Indels 0; Gaps 0;
XX
QY 1 MITIVLILAYLGSIPSGLWIGOVFFQINLREHSGSGTGTNTFRILGKKGAMATFVID 60
DB 28 MKLLFTTIAYLLGSIPTGLWIGQYFVHINLREHSGSGTGTNTFRILGKAGTATLAD 87
XX
QY 61 FFKGTLATLPIIFHLQGVSPFLIFGLLAVIGHTFPFAGFKGKAVATSAGVIFGPAPIF 120
DB 88 MFKGTLILPIIFGWTSSISAIGFPFVLGHTFPFIPANFKGKAVATSAGVILGPAPLY 147
XX
QY 121 CLYLAIFFGALYGLSMISLSSVTASTAAVIGVLPLFLGFLFSLNSDLSFIALIALASL 180
DB 148 LFFLASIFVLVLYLFSMISLASVSAVIGVLSVLTPPAIHFLPNVDYFLTFVILLAFI 207
XX
QY 181 IIRHKDNIAIRIKNKTENLVPMGLNLTHODPKK 213
DB 208 IIRHKDNISRIKHHTENLVPMGLNLKQVPPK 240
XX
RESULT 6
ABP26818
ID ABP26818 standard; Protein: 212 AA.
XX
XX AC ABP26818;
XX
XX 02-JUL-2002 (first entry)
XX
```

```
XX Streptococcus polypeptide SEQ ID NO 2812.
DE
XX
XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW group A streptococcus; Streptococcus pyogenes; antibacterial;
KW antiinflammatory; infection; vaccine; meningitis; gene therapy.
XX
XX Streptococcus agalactiae.
OS
XX
XX WO200234771-A2.
PN
XX
XX 02-MAY-2002.
PD
XX
XX 29-OCT-2001; 2001WO-GB04789.
XX
XX 27-OCT-2000; 2000GB-0026333.
PR 24-NOV-2000; 2000GB-0028727.
PR 07-MAR-2001; 2001GB-0005640.
XX
XX (CHIR-) CHIRON SPA.
PA (GENO-) INST GENOMIC RES.
XX
XX Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;
PI Tettelin H;
XX
XX WPI: 2002-352536/38.
DR N-PSDB; ABN67449.
XX
XX New Streptococcus protein for the treatment or prevention of infection
PT or disease caused by Streptococcus bacteria, such as meningitis, and
PT for detecting a compound that binds to the protein -
XX
XX Claim 1; Page 3432; 4525pp; English.
XX
XX The invention relates to a protein (ABP25413-ABP30895) from group B
CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC the specification. The proteins have antibacterial and antiinflammatory
CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC antibodies that bind (I) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC Nucleic acids encoding (I) are used to detect Streptococcus in a
CC biological sample. (I) is used to determine whether a compound binds to
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC used as a vaccine or diagnostic composition. The disease caused by
CC Streptococcus that is prevented or treated may be meningitis. Nucleic
CC acid encoding (I) may be used to recombinantly produce (I) and may be
CC used in gene therapy. Antibodies to (I) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
CC Streptococcus proteins.
XX
XX Sequence 212 AA;
XX
XX Query Match 66.1%; Score 708; DB 23; Length 212;
XX Best Local Similarity 63.2%; Pred. No. 2.6e-74;
XX Matches 134; Conservative 33; Mismatches 45; Indels 0; Gaps 0;
XX
QY 1 MITIVLILAYLGSIPSGLWIGOVFFQINLREHSGSGTGTNTFRILGKKGAMATFVID 60
DB 1 MNIIIMIIAYLLGSIPTGLWIKYFYQVNLROHSGSGTGTNTFRILGKAGIVTLTID 60
XX
QY 61 FFKGTLATLPIIFHLQGVSPFLIFGLLAVIGHTFPFAGFKGKAVATSAGVIFGPAPIF 120
DB 61 ILKGTLATLPIILGITTVSPFFIGFFAIIGHTFPFIFAQFKGKAVATSAGVILGPAPSF 120
XX
QY 121 CLYLAIFFGALYGLSMISLSSVTASTAAVIGVLLPLFLGFLFSLNSDLSFIALIALASL 180
DB 121 FLYLLVFLTLVLYLFSMISLSSITVAVVGILSVLIFPLVGLFILDYDWTFTVILMALT 180
XX
QY 181 IIRHKDNIAIRIKNKTENLVPMGLNLTHODPK 212
DB 181 IIRHKDNIAIRIKRQENLVPMGLNLKQKNK 212
XX
```


Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	548	51.2	114	4	US-09-222-938A-13	Sequence 13, Appl
2	468	43.7	193	4	US-09-222-938A-70	Sequence 70, Appl
3	463.5	43.3	204	4	US-09-134-001C-5239	Sequence 5239, Ap
4	101	9.4	347	4	US-09-134-001C-3960	Sequence 3960, Ap
5	93	8.7	323	4	US-09-134-001C-4635	Sequence 4635, Ap
6	92.5	8.6	405	4	US-09-134-001C-4999	Sequence 4999, Ap
7	91.5	8.5	287	4	US-09-134-001C-5055	Sequence 5055, Ap
8	91.5	8.5	438	2	US-08-677-049-9	Sequence 9, Appl
9	91.5	8.5	518	4	US-09-134-001C-4744	Sequence 4744, Ap
10	89.5	8.4	400	4	US-09-134-001C-2912	Sequence 2912, Ap
11	88	8.2	454	4	US-09-134-001C-3501	Sequence 3501, Ap
12	87.5	8.2	799	4	US-09-165-396-4	Sequence 4, Appl
13	86	8.0	655	4	US-09-245-808-1	Sequence 1, Appl
14	86	8.0	901	4	US-09-134-001C-5389	Sequence 5389, Ap
15	85.5	8.0	190	4	US-09-134-001C-5000	Sequence 5000, Ap
16	85.5	8.0	460	2	US-08-677-049-10	Sequence 10, Appl
17	85.5	8.0	635	2	US-09-014-969-11	Sequence 11, Appl
18	85	7.9	237	2	US-08-849-480A-7	Sequence 7, Appl
19	85	7.9	299	2	US-09-097-759-3	Sequence 3, Appl
20	85	7.9	306	2	US-09-097-759-2	Sequence 2, Appl
21	85	7.9	315	4	US-09-134-001C-5446	Sequence 5446, Ap
22	85	7.9	668	4	US-09-134-001C-3430	Sequence 3430, Ap
23	84.5	7.9	229	4	US-09-227-357-237	Sequence 237, Appl
24	84.5	7.9	530	4	US-09-134-001C-4510	Sequence 4510, Ap
25	84	7.8	372	2	US-08-626-685A-8	Sequence 8, Appl
26	84	7.8	372	4	US-08-993-088A-2	Sequence 2, Appl
27	84	7.8	372	4	US-08-993-424B-2	Sequence 2, Appl

```

RESULT 2
US-09-222-938A-70
: Sequence 70. Application US/09222938A
: Patent No. 6437108
: GENERAL INFORMATION:
: APPLICANT: Youngman, Philip
: APPLICANT: Fritz, Christian
: APPLICANT: Murphy, Christopher
: APPLICANT: Guzman, Luz-Maria
: TITLE OF INVENTION: ESSENTIAL BACTERIA
: FILE REFERENCE: 07334/060001
: CURRENT APPLICATION NUMBER: US/09/22
: CURRENT FILING DATE: 1998-12-30
: NUMBER OF SEQ ID NOS: 102
: SEQUENCE: FastSeq for Windows Version 1.0
: SEQ ID NO 70

```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 26, 2002, 00:31:11 : Search time 332.091 Seconds
(without alignments)
61.660 Million cell updates/sec

Title: US-10-068-080-1
Perfect score: 1071
Sequence: 1 MITIVLLILAYLLGSLPSGL.....NKTNLVPWGLNLTHQDPKK 213

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1071	100.0	213	2 C95098	conserved hypothet
2	1071	100.0	213	2 C97966	conserved hypothet
3	716.5	66.9	213	2 B86747	conserved hypothet
4	468	43.7	193	1 A69892	conserved hypothet
5	463.5	43.3	198	2 AD1235	B. subtilis YneS p
6	459.5	42.9	198	2 AB1598	conserved hypothet
7	441.5	41.2	202	2 A89911	conserved hypothet
8	332	31.0	192	2 E70359	conserved hypothet
9	300.5	28.1	222	1 S75137	hypothetical prote
10	298	27.8	206	2 G83854	hypothetical prote
11	291.5	27.2	198	2 A75324	conserved hypothet
12	282.5	26.4	201	2 AC3593	probable integral
13	282	26.3	202	2 AB2737	conserved hypothet
14	282	26.3	205	2 H97517	hypothetical prote
15	275.5	25.7	198	2 C97208	uncharacterized co
16	275.5	25.7	200	2 C81894	probable integral
17	274	25.6	208	2 B82313	conserved hypothet
18	269.5	25.2	196	2 C72253	conserved hypothet
19	266.5	24.9	203	2 AF0892	probable membrane
20	266	24.8	200	2 H81126	conserved hypothet
21	265.5	24.8	216	2 AB0080	probable membrane
22	263.5	24.6	205	1 A65094	ygiH protein - Esc
23	263.5	24.6	205	2 F91121	hypothetical prote
24	263.5	24.6	205	2 D81378	hypothetical prote
25	256.5	23.9	202	2 D81378	probable integral
26	255.5	23.9	226	2 AC1868	hypothetical prote
27	244.5	22.8	199	2 E64146	hypothetical prote
28	239	22.3	189	2 A83573	conserved hypothet
29	236.5	22.1	257	2 A82894	conserved hypothet

30	236	22.0	218	2 E87552	conserved hypothet
31	227	21.2	224	2 G90548	conserved hypothet
32	223.5	20.9	239	1 C64227	hypothetical prote
33	214.5	20.0	239	2 S73812	hypothetical prote
34	213	19.9	220	2 B71811	hypothetical prote
35	211	19.7	262	1 E64708	conserved hypothet
36	159.5	14.9	203	2 F72273	conserved hypothet
37	156	14.6	559	2 B75477	conserved hypothet
38	144	13.4	195	2 A75295	conserved hypothet
39	122.5	11.4	963	2 AF2119	hypothetical prote
40	113.5	10.6	348	2 T12290	NADH2 dehydrogenas
41	109	10.2	504	2 AB0839	probable membrane
42	107.5	10.0	405	2 B69200	hypothetical prote
43	107	10.0	640	2 D90174	hypothetical prote
44	106	9.9	267	2 AI2794	conserved hypothet
45	106	9.9	310	2 H97573	hypothetical prote

ALIGNMENTS

RESULT 1

C95098
conserved hypothetical protein SP0851 [imported] - Streptococcus pneumoniae (strai
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
C:Accession: C95098
R:Rettelin, H.; Neilson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtz
son, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Mo
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoni
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: C95098
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-213 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK74980.1; PID:g14972323; GSPDB:GN00164; T1
A:Experimental source: strain TIGR4
C:Genetics:
C:Gene: SP0851
C:Superfamily: Escherichia coli ygiH protein

Query Match	100.0%	Score 1071;	DB 2;	Length 213;
Best Local Similarity	100.0%;	Pred. No. 1.5e-80;		
Matches 213;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY 1	MITIVLLILAYLLGSLPSGLWIGOVFFQINLRHSGSGNTGTTNTRILKKGKAGMATFVID 60			
Db 1	MITIVLLILAYLLGSLPSGLWIGOVFFQINLRHSGSGNTGTTNTRILKKGKAGMATFVID 60			
QY 61	FFKGTATLPIIFHLOGVSPLIIFGLAVIHTFPIFAGKKGAVATSAGVIFGPAIF 120			
Db 61	FFKGTATLPIIFHLOGVSPLIIFGLAVIHTFPIFAGKKGAVATSAGVIFGPAIF 120			
QY 121	CLYLAIIFGALYLSMISLSSVTASIAAIVGLVLPGLFGLSNDLSFIATILASL 180			
Db 121	CLYLAIIFGALYLSMISLSSVTASIAAIVGLVLPGLFGLSNDLSFIATILASL 180			
QY 181	IIIRHKDNTARIKNKTENLVPWGLNLTHQDPKK 213			
Db 181	IIIRHKDNTARIKNKTENLVPWGLNLTHQDPKK 213			

RESULT 2

C97966
conserved hypothetical protein spr0755 [imported] - Streptococcus pneumoniae (str
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, E
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren,

Db 181 IILRHRTNLKRINKNCESLVPGLNLSKOKEX 213

RESULT 4
A69892
Conserved hypothetical protein yneS - Bacillus subtilis
C/Species: Bacillus subtilis
C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C/Accession: A69892
R/Kunst, F.; Ogawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.;
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrar
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.;
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulle
Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lard
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Port
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Sc
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.;
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uch
t, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yosh
A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus sub
A:Reference number: A69580; MUID:98044033; PMID:9384377
A:Accession: A69892
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-193 <KUN>
A:Cross-references: GB:299113; GB:AL009126; NID:g2634090; PIDN:CAB13690.1; PID:g26
A:Experimental source: strain 168
C/Genetics:
A:Gene: yneS
C:Superfamily: Escherichia coli ygiH protein

Query Match 43.7%; Score 468; DB 1; Length 193;
Best Local Similarity 48.0%; Pred. No. 2.7e-31;
Matches 97; Conservative 32; Mismatches 63; Indels 10; Gaps 3;
QY 1 MITVLILAYLLGSIPLGWLIGVVFQINLRHSGSGTGTNTFRILGKKAGMATEFVD 60
Db 1 MIALLIILAYLIGSIFGLVIGLAKGIDIREHGSGNLGATNATFRILGVKAGSVIATGD 60
QY 61 FFKGTALTLLPIIFHLQGVSLPIFGLLAVIGHTPPIFAGFKGKAVATSAGVIFGAPIF 120
Db 61 ILKGTALATPLMHVD-IHPLLAGVFAVLGHVFPFPAKFKGKAVATSGGVLLFYAPLL 119
QY 121 CLYLAIIFGALYGLSMISLSSVTASTANIGVLLFPFLFGFILSNYDSLFATILASL 180
Db 120 FITWVAFPIFYLTKEVLSLSMLTGTIVT-----YSFV--HDTYLLIVVLTITF 170
QY 181 IIRHKDNIARIKNTNLVNPW 202
Db 171 VIYHRANKRIINKTEPKVKW 192

RESULT 5
AD1235
B. subtilis yneS protein homolog lmol284 [imported] - Listeria monocytogenes (stra
C/Species: Listeria monocytogenes
C/Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C/Accession: AD1235
R/Glaser, P.; Grangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bl
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsi
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam,
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; We
A:Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AD1235
A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-198 <GLA>

A:Cross-references: GB:NC_003210; PIDN:CAC99362.1; PID:g16410700; GSPDB:GN00177

A:Experimental source: strain EGD-e

C:Genetics:

A:Gene: lml284

C:Superfamily: Escherichia coli ygiH protein

Query Match 43.3%; Score 463.5; DB 2; Length 198;

Best Local Similarity 47.8%; Pred. No. 6.4e-31;

Matches 96; Conservative 38; Mismatches 56; Indels 11; Gaps 4;

QY 4 IVLLILAYLGSPSGLWIGVFFQINLREHSGNGTNTTFRILKKGKAGMATFVIDFFK 63

DB 6 ILLSLAYVIGSPSGLWIGKIFPKKIDREFGSGNLGATNSFRVLKAGSIVTVMDILK 65

QY 64 GTLTLPLPIHFHQLGSP--LIFGLLAVIGHTFPFAGFKGKAVATSAGVIFGFAPIFC 121

DB 66 GTVATLLPFFQLNVHDFWLLTGAFATIGHSPFAGFRGKAVATSAGVILAYAPLLF 125

QY 122 LYLAIIFFGALYLGSMISLSSVTASTAAVIGVLLPFLFGFILSNYDSLFATILALASLI 181

DB 126 VAAALVFLVTLKLSKYVSLSMMIGALAAI--ISLF--MG-----DWILLIVACIALFV 176

QY 182 IIRHKDNIARIKNTENLVWP 202

DB 177 IWRHRANITRIRNGEPEPKIKW 197

RESULT 6

AB1598

conserved hypothetical protein, B. subtilis yneS protein homolog lin1323 [imported] - Li

C:Species: Listeria innocua

C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001

C:Accession: AB1598

R:Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker

; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.

D.; Jones, L.M.; Karst, U.

Science 294, 849-852, 2001

A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma

ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wenland,

A:Title: Comparative genomics of Listeria species.

A:Reference number: AB1077; MUID:21537279; PMID:11679669.

A:Accession: AB1598

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-198 <GLA>

A:Cross-references: GB:AL592022; PIDN:CAC96554.1; PID:g16413796; GSPDB:GN00178

A:Experimental source: strain Clp11262

C:Genetics:

A:Gene: lin1323

C:Superfamily: Escherichia coli ygiH protein

Query Match 42.9%; Score 459.5; DB 2; Length 198;

Best Local Similarity 46.0%; Pred. No. 1.4e-30;

Matches 93; Conservative 41; Mismatches 55; Indels 13; Gaps 4;

QY 4 IVLLILAYLGSPSGLWIGVFFQINLREHSGNGTNTTFRILKKGKAGMATFVIDFFK 63

DB 6 ILLSLAYVIGSPSGLWIGKIFPKKIDREFGSGNLGATNSFRVLKAGSIVTVMDILK 65

QY 64 GTLTLPLPIHFHQLGSP--LIFGLLAVIGHTFPFAGFKGKAVATSAGVIFGFAPIFC 120

DB 66 GTVATLLPFFQLNVHDFWLLTGAFATIGHSPFAGFRGKAVATSAGVILAYAPLL 124

QY 121 CLYLAIFFGALYLGSMISLSSVTASTAAVIGVLLPFLFGFILSNYDSLFATILALASL 180

DB 125 FVAALVFLVTLKISKYVSLSMMIGALAAI-----ISFFMG--DWILLIVACIALF 175

QY 181 IIRHKDNIARIKNTENLVWP 202

DB 176 VIWRHRANITRIRNGEPEPKIKW 197

RESULT 7

AB9911

conserved hypothetical protein Sall187 [imported] - Staphylococcus aureus (strain N

C:Species: Staphylococcus aureus

C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001

C:Accession: AB9911

R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.

ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu

C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiratsugu, K.

Lancet 357, 1225-1240, 2001

A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.

A:Reference number: AB9758; MUID:21311952; PMID:11418146

A:Accession: AB9911

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-202 <KUR>

A:Cross-references: GB:BA000018; PID:g13701150; PIDN:BA042445.1; GSPDB:GN00149

A:Experimental source: strain N315

C:Genetics:

A:Gene: Sall187

C:Superfamily: Escherichia coli ygiH protein

Query Match 41.2%; Score 441.5; DB 2; Length 202;

Best Local Similarity 42.9%; Pred. No. 4.1e-29;

Matches 90; Conservative 43; Mismatches 60; Indels 17; Gaps 3;

QY 1 MITIVLLIAYLGSPSGLWIGVFFQINLREHSGNGTNTTFRILKKGKAGMATFVID 60

DB 1 MMIIIVMLLSYLTGAPPSGVICKLFKKDIRQFGSGNTGATNSFRVLGRPAGFLVTFD 60

QY 61 FPKGTATLPLPIHFHQLGSP-----LIFGLLAVIGHTFPFAGFKGKAVATSAGV 112

DB 61 IFKGFITVFPPMLPWHADGPSTFTNGLIVGLFALGHVFPVILKFGGKAVATSAGV 120

QY 113 IFESFAPICFLYLAIFFGALYLGSMISLSSVTASTAAVIGVLLPFLFGFILSNYDSLFIA 172

DB 121 VLGVNPILLILALAIFFIVLKFYKVSASIVAAICCVIGSL-----IIQDYILLVVS 173

QY 173 IILASLIIIRHKDNIARIKNTENLVWP 202

DB 174 FLVSI--ILIIHRHSNIARIFRGEPEPKIKW 201

RESULT 8

E70359

conserved hypothetical protein aq_676 - Aquifex aeolicus

C:Species: Aquifex aeolicus

C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 20-Sep-1999

C:Accession: E70359

R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.

V.

Nature 392, 353-358, 1998

A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.

A:Reference number: A70300; MUID:98196666; PMID:9537320

A:Accession: E70359

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-192 <AQ>

A:Cross-references: GB:AE000701; NID:g2983260; PIDN:AAC06869.1; PID:g2983275; GB:A

A:Experimental source: strain VF5

C:Genetics:

A:Gene: aq_676

C:Superfamily: Escherichia coli ygiH protein

Query Match 31.0%; Score 332; DB 2; Length 192;

Best Local Similarity 39.8%; Pred. No. 3.5e-20;

Matches 80; Conservative 39; Mismatches 68; Indels 14; Gaps 6;

QY 1 MITIVLLIAYLGSPSGLWIGVFFQINLREHSGNGTNTTFRILKKGKAGMATFVID 60

DB 1 MKALFLVIFAYLLGSITTFGEVIAKL-KGVDLRNWGSNGVATNVTALGKKYGLVFFLD 59

QY 61 FPKGTATLPLPIHFHQLGSPVPLIFGLLAVIGHTFPFAGFKGKAVATSAGVIFGFAP 119

Db 60 FLKGFIPALIAVSKSGIDSWLFTGLASVLGHMYPVFFGKGGKGVATAGLVVFAVSPS 119
 QY 120 FCYLAIIFPGALYLGSMISLSSVTASIAVIGVLLPFLFGFILSN--DSLEIAIILAL 177
 Db 120 VALFSLVWLGIFLWKRYVSLASITATISAF-----LFLFV-AGYPVNVLFMAIV--I 169
 QY 178 ASLIIIRHKDNIARIKNKTEN 198
 Db 170 GALLIYHRENIINRLTGREH 190

RESULT 9
 S75137
 hypothetical protein sll1973 - *Synechocystis* sp. (strain PCC 6803)
 C:Species: *Synechocystis* sp.
 A:Variety: PCC 6803
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
 R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
 DNA Res. 3, 109-136, 1996
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*
 S.
 A:Reference number: S74322; MUID:97061201; PMID:8905231
 A:Accession: S75137
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-222 <KAN>
 A:Cross-references: EMBL:D90910; GB:AB001339; NID:g1652956; PIDN:BAAL7999.1; PID:g165308
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
 C:Superfamily: *Escherichia coli* ygiH protein

Query Match 28.1%; Score 300.5; DB 1; Length 222;
 Best Local Similarity 34.2%; Pred. No. 1.5e-17;
 Matches 75; Conservative 37; Mismatches 66; Indels 41; Gaps 6;

QY 1 MITIVILLIAYLGSPGLWICQVFFQINLREHSGNTGTTNFRILGKKKAGMATFVID 60
 Db 6 LILCCLLLIYLGSIPTGYLAGKLLGIDIREHSGKSTGATNVRTLGKPAIAVLAD 65
 QY 61 FFKGTATLPIIFHLQVSPLI-----FGLLAVIGHTFPFAGFKGKAVATS 109
 Db 66 ISKGVMAVALVRAIYSDWLPALPAWQNLTLGVAIVLVGHKSIFLKFSGKSVATS 125
 QY 110 AGVIFGAPIFCULYLAIF-----IFGALYLGSMISLSSVTASIAVIGVLL-----FPL 158
 Db 126 LGVLF-----MLNIWLALGTALTELTVFTRIVSLSSIVAAL-AVNGIALALQPPYLA 180
 QY 159 FGILSNYDSLFTAILALASLIIIRHKDNIARIKNKTE 197
 Db 181 FFLAGMY-----VIVRHTNRILOQTE 205

RESULT 10
 G83854
 hypothetical protein BH1639 [imported] - *Bacillus halodurans* (strain C-125)
 C:Species: *Bacillus halodurans*
 C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
 C:Accession: G83854
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, F.; Hira
 Nucleic Acids Res. 28, 4317-4331, 2000
 A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and
 A:Reference number: A83650; MUID:20512582; PMID:11058132
 A:Accession: G83854
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-206 <STO>
 A:Cross-references: GB:AP001512; GB:BA000004; NID:g10174030; PIDN:BA805358.1; GSPDB:GN0
 A:Experimental source: strain C-125
 C:Genetics:
 A:Gene: BH1639
 C:Superfamily: *Escherichia coli* ygiH protein

Query Match 27.8%; Score 298; DB 2; Length 206;
 Best Local Similarity 33.0%; Pred. No. 2.2e-17;
 Matches 65; Conservative 49; Mismatches 75; Indels 8; Gaps 4;

QY 5 VLLIALLGSPGLWICQVFFQINLREHSGNTGTTNFRILGKKKAGMATFVIDFFKG 64
 Db 6 LIIIGSYLGSVSFSYIIAKIKKVDIROHSGNAGATNTRLVVGPGPAVTIVLLDLKG 65
 QY 65 TIATLLPIIFHLQVSP--PLIFGLLAVIGHTFPFAGFKGKAVATSAGVIFGFAPIFCL 122
 Db 66 VIAVVTVQLTPDGDGFWFAAAAGIAAIGHNWIYVYGFGRGKVATTIGVLSLVPLAAV 125
 QY 123 YLAIIFGALYLGSMISLSSVT-ASIAVIGVLLFPLFGFILSNYDSLFTAILALASLI 181
 Db 126 LAGVIAIGSIVWTRYVSLGSLLEVTALLVAVLQWFGYPVA---YIVLTIIVAILSM- 181
 QY 182 IIRHKDNIARIKNKTEN 198
 Db 182 -WHRNSIORLLSGTEN 197

RESULT 11
 A75324
 conserved hypothetical protein - *Deinococcus radiodurans* (strain R1)
 C:Species: *Deinococcus radiodurans*
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
 C:Accession: A75324
 R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson,
 M.; Shen, M.; Vanathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski,
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
 Science 286, 1571-1577, 1999
 A:Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R
 A:Reference number: A75250; MUID:20036896; PMID:10567266
 A:Accession: A75324
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-198 <WHI>
 A:Cross-references: GB:AE002039; GB:AE000513; NID:g6459810; PIDN:AAF11571.1; PID:g
 A:Experimental source: strain R1
 C:Genetics:
 A:Gene: DR2021
 A:Map position: 1
 C:Superfamily: *Escherichia coli* ygiH protein

Query Match 27.2%; Score 291.5; DB 2; Length 198;
 Best Local Similarity 34.5%; Pred. No. 7.4e-17;
 Matches 68; Conservative 39; Mismatches 81; Indels 9; Gaps 3;

QY 1 MITIVILLIAYLGSPGLWICQVFFQINLREHSGNTGTTNFRILGKKKAGMATFVID 60
 Db 3 LTALLALLSYLIGAIIPAAAWLARA-RGVDIRKVGSGNSGATNVLRLSLGKGPALLVASED 61
 QY 61 FFKGTATLPIIFHLQVSPLIIFGLLAVIGHTFPFAGFKGKAVATSAGVIFGFAPIF 120
 Db 62 ILKGVLAVALRALGALSAEWAALCGVLAVIGHNFSPLAFRGKGVATSGVIAIDLPVL 121
 QY 121 CLYLAIIFGALYLGSMISLSSVTASIAVIGVLLFPLFGFILSNYDSLFTAILALASL 180
 Db 122 GLTFTVLAACWMLTRFVSAGSIMGAFIAGALVVLVLP-----RPTWDR---AAVFLAAL 173
 QY 181 IIRHKDNIARIKNKTE 197
 Db 174 LVVQHRNIRKIQAGTE 190

RESULT 12
 AC3593
 Probable integral membrane protein [imported] - *Brucella melitensis* (strain 16M)
 C:Species: *Brucella melitensis*
 C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 15-Feb-2002
 C:Accession: AC3593
 R:DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mujeer, C.; Los, T.; Iva

.. Mazur, M.; Coltsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen *Brucella melitens*
A:Reference number: AB3252; PMID:11756688
A:Accession: AC3593
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-201 <KUR>
A:Cross-references: GB:AE008918; PIDN:AAL53910.1; PID:gl7984852; GSPDB:GN00191
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BME110668
A:Map position: 11
C:Superfamily: *Escherichia coli* ygiH protein

Query Match 26.4%; Score 282.5; DB 2; Length 201;
Best Local Similarity 36.6%; Pred. No. 4.5e-16;
Matches 71; Conservative 31; Mismatches 81; Indels 11; Gaps 3;

QY 7 LILAYLLGSPGLWIGVFFQINLRHSGNVTGTTNFRILGKKAGMATFVIDFFKGTLL 66
DB 15 LIFGYVLGSPGLILTRAGLDVRAIGSGNIGATNLTGNKKLAATAATLLDLKGTAA 74
QY 67 ATLLPIIFHLOGVSPILFGLLAVIGHTFPFAGFGKGVAVTSAGVIFGFAPIFCLYLAI 126
DB 75 AAL--TAHFGQNAATAGGAFIGHLPFWIGFGKGVATYGLVGLAWAGALVFPA 132
QY 127 IFFGALYLGSMISLSSVTASIAAIGVLLFPFAGFGKGVAVTSAGVIFGFAPIFCLYLAI 186
DB 133 AWIVTALLARYSSLSALVASL--VVPYALY-----SRGNQALAAFAIMTVIVFIKHR 183
QY 187 DNIARIKNKTENLV 200
DB 184 ANISRLNGTESKI 197

RESULT 13
AB2737
conserved hypothetical protein Atul306 [imported] - *Agrobacterium tumefaciens* (strain C58)
C:Species: *Agrobacterium tumefaciens*
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 01-Feb-2002
C:Accession: AB2737
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer *Agrobacterium tumefaciens* C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AB2737
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-202 <KUR>
A:Cross-references: GB:AE008688; PIDN:AAL42312.1; PID:gl7739714; GSPDB:GN00186
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atul306
A:Map position: circular chromosome
C:Superfamily: *Escherichia coli* ygiH protein

Query Match 26.3%; Score 282; DB 2; Length 202;
Best Local Similarity 35.1%; Pred. No. 4.5e-16;
Matches 68; Conservative 40; Mismatches 74; Indels 12; Gaps 4;

QY 8 ILAYLLGSPGLWIGVFFQINLRHSGNVTGTTNFRILGKKAGMATFVIDFFKGTLL 67
DB 16 LIGYLLGSPGLILTRAGLDVVRKIGSGNIGATNLTGNKKLAATAATLLDLKGTAA 75
QY 68 TLLPIIFHLOGV-SPLIFGLLAVIGHTFPFAGFGKGVAVTSAGVIFGFAPIFCLYLAI 126
DB 76 VL--VANALMGYEASLVAGFFAFLGHLFPVWLFGKGVAVYIGVLLGAAPLMLAFAL 133

QY 127 IFFGALYLGSMISLSSVTASIAAIGVLLFPFAGFGKGVAVTSAGVIFGFAPIFCLYLAI 186
DB 134 IWLATAFIYRYSLS-----ALLAMLIIPVALVGLPEKTMALVTLSSVISM--WKHR 184
QY 187 DNIARIKNKTENLV 200
DB 185 ENIRRLMAGTESRI 198

RESULT 14
H97517
hypothetical protein AGR_C_2402 [imported] - *Agrobacterium tumefaciens* (strain C58)
C:Species: *Agrobacterium tumefaciens*
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
C:Accession: H97517
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Go
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markel
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent *Agrobacteri*
A:Reference number: A97359; PMID:11743194
A:Accession: H97517
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-205 <KUR>
A:Cross-references: GB:AE007869; PIDN:AAK87097.1; PID:gl5156359; GSPDB:GN00169
C:Genetics:
A:Gene: AGR_C_2402
A:Map position: circular chromosome
C:Superfamily: *Escherichia coli* ygiH protein

Query Match 26.3%; Score 282; DB 2; Length 205;
Best Local Similarity 35.1%; Pred. No. 4.5e-16;
Matches 58; Conservative 40; Mismatches 74; Indels 12; Gaps 4;

QY 8 ILAYLLGSPGLWIGVFFQINLRHSGNVTGTTNFRILGKKAGMATFVIDFFKGTLL 67
DB 19 LIGYLLGSPGLILTRAGLDVVRKIGSGNIGATNLTGNKKLAATAATLLDLKGTAA 78
QY 68 TLLPIIFHLOGV-SPLIFGLLAVIGHTFPFAGFGKGVAVTSAGVIFGFAPIFCLYLAI 126
DB 79 VL--VANALMGYEASLVAGFFAFLGHLFPVWLFGKGVAVYIGVLLGAAPLMLAFAL 136
QY 127 IFFGALYLGSMISLSSVTASIAAIGVLLFPFAGFGKGVAVTSAGVIFGFAPIFCLYLAI 186
DB 137 IWLATAFIYRYSLS-----ALLAMLIIPVALVGLPEKTMALVTLSSVISM--WKHR 187
QY 187 DNIARIKNKTENLV 200
DB 188 ENIRRLMAGTESRI 201

RESULT 15
C97208
uncharacterized conserved membrane protein, YgiH/UPF0078 family CAC2500 [imported]
C:Species: *Clostridium acetobutylicum*
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001
C:Accession: C97208
R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacteri
A:Reference number: A96900; PMID:21359325; PMID:21359325
A:Accession: C97208
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-198 <KUR>
A:Cross-references: GB:AE001437; PIDN:AAK80454.1; PID:gl5025522; GSPDB:GN00168
A:Experimental source: *Clostridium acetobutylicum* ATCC824
C:Genetics:
A:Gene: CAC2500
C:Superfamily: *Escherichia coli* ygiH protein

Query Match 25.7%; Score 275.5; DB 2; Length 198;

Best Local Similarity 32.0%; Pred. No. 1.5e-15;
 Matches 66; Conservative 39; Mismatches 82; Indels 19; Gaps 4;

QY	1	MITIVILLIAYLLGSLIPSGLNICQVFFQINLREHSGNLTGTTNTFRILCKKAGMATFVID	60
Db	2	IIVITVMSVFLGSGIPTGYLITKLSGIDVRTKSGNIGSTNVKRVAGTKISMITQVMD	61
QY	61	FPKGTLATLLPIIFHLQGVSP-----IFGLLAVIGHTPEIFAGFKGKAVATSAGVI	113
Db	62	ILKGIIPVLLCMLIASKIKLPISTSMYLSIIIVIAVLGHDTPTFLGNGGKGVNTTVGAF	121
QY	114	FGPAIFCULYLAIFFGALYLGSMISLSSVTASIAAIG--VLLFPLFGFILSNYDSLFI	171
Db	122	FLAPAAVLGAVVYFVLRFTKIVSIKSIAGITMPACIALRLPI-----EITV	172
QY	172	AIILALASLIIRHKONIARIKNKTE	197
Db	173	CAIIA-CGLLIIRHKDNLRLVNEE	197

Search completed: December 26, 2002, 01:27:05
 Job time : 333.091 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 26, 2002, 01:16:25 ; Search time 89.1872 Seconds
(without alignments)
41.350 Million cell updates/sec

Title: US-10-068-080-1
Perfect score: 1071
Sequence: 1 MITVILLIAYLGSIPLSGL.....NKTENLVPLGNLTHQDPKK 213

Scoring table: BLOSUM62

Searched: Gap0 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters: 109717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

- Database : Published Applications AA:*
- 1: /cgn2.6/ptodata/2/pubpaa/US08_NEW_PUB.pap.*
 - 2: /cgn2.6/ptodata/2/pubpaa/PCT_NEW_PUB.pap.*
 - 3: /cgn2.6/ptodata/2/pubpaa/US06_NEW_PUB.pap.*
 - 4: /cgn2.6/ptodata/2/pubpaa/US06_PUBCOMB.pap.*
 - 5: /cgn2.6/ptodata/2/pubpaa/US07_NEW_PUB.pap.*
 - 6: /cgn2.6/ptodata/2/pubpaa/US07_PUBCOMB.pap.*
 - 7: /cgn2.6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pap.*
 - 8: /cgn2.6/ptodata/2/pubpaa/US08_PUBCOMB.pap.*
 - 9: /cgn2.6/ptodata/2/pubpaa/US09_NEW_PUB.pap.*
 - 10: /cgn2.6/ptodata/2/pubpaa/US09_PUBCOMB.pap.*
 - 11: /cgn2.6/ptodata/2/pubpaa/US10_NEW_PUB.pap.*
 - 12: /cgn2.6/ptodata/2/pubpaa/US10_PUBCOMB.pap.*
 - 13: /cgn2.6/ptodata/2/pubpaa/US60_NEW_PUB.pap.*
 - 14: /cgn2.6/ptodata/2/pubpaa/US60_PUBCOMB.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1071	100.0	213	10	US-09-815-242-13329 Sequence 13329, A
2	1071	100.0	213	10	US-09-815-242-13536 Sequence 13536, A
3	1071	100.0	213	12	US-10-068-080-1 Sequence 1, Appli
4	468	43.7	193	12	US-10-068-080-3 Sequence 3, Appli
5	443.5	41.4	202	10	US-09-823-246-2 Sequence 2, Appli
6	438.5	40.9	202	10	US-09-815-242-5305 Sequence 5305, Ap
7	438.5	40.9	202	10	US-09-815-242-12137 Sequence 12137, A
8	266.5	24.9	203	10	US-09-815-242-13757 Sequence 13757, A
9	263.5	24.6	205	10	US-09-815-242-10303 Sequence 10303, A
10	252.5	23.6	205	10	US-09-815-242-11713 Sequence 11713, A
11	244.5	22.8	199	10	US-09-815-242-10997 Sequence 10997, A
12	239	22.3	189	10	US-09-815-242-11799 Sequence 11799, A
13	213	19.9	220	10	US-09-815-242-11614 Sequence 11614, A
14	211	19.7	252	10	US-09-815-242-11453 Sequence 11453, A
15	105.5	9.9	434	10	US-09-778-927A-79 Sequence 79, Appli
16	102	9.5	650	10	US-09-815-242-13341 Sequence 13341, A
17	100.5	9.4	456	10	US-09-795-693-2 Sequence 2, Appli
18	97.5	9.1	466	10	US-09-778-927A-78 Sequence 78, Appli
19	97.5	9.1	466	10	US-09-778-927A-81 Sequence 81, Appli

20	95.5	8.9	337	10	US-09-815-242-11246 Sequence 11246, A
21	95	8.9	769	10	US-09-815-242-10575 Sequence 10575, A
22	95	8.9	782	10	US-09-815-242-4893 Sequence 4893, Ap
23	94.5	8.8	382	10	US-09-795-693-23 Sequence 23, Appli
24	93	8.7	452	10	US-09-815-242-13983 Sequence 13983, A
25	90.5	8.5	220	9	US-09-860-670-129 Sequence 129, App
26	90.5	8.5	220	9	US-09-764-868-949 Sequence 949, App
27	90.5	8.5	220	9	US-09-764-868-1231 Sequence 1231, Ap
28	88.5	8.3	428	10	US-09-741-869-394 Sequence 394, App
29	88.5	8.3	1594	10	US-09-971-121-4 Sequence 4, Appli
30	88.5	8.3	1642	10	US-09-971-121-2 Sequence 2, Appli
31	88.5	8.3	1642	12	US-10-090-458-5 Sequence 5, Appli
32	88	8.2	760	10	US-09-833-017-26 Sequence 26, Appli
33	87	8.1	264	10	US-09-815-242-13439 Sequence 13439, A
34	87	8.1	264	10	US-09-815-242-13555 Sequence 13555, A
35	87	8.1	1624	12	US-10-090-454-2 Sequence 2, Appli
36	86.5	8.1	246	10	US-09-815-242-11377 Sequence 11377, A
37	86.5	8.1	418	10	US-09-815-242-12003 Sequence 12003, A
38	86.5	8.1	453	10	US-09-815-242-10872 Sequence 10872, A
39	85.5	8.0	448	10	US-09-815-242-12421 Sequence 12421, A
40	85.5	8.0	535	10	US-09-795-693-20 Sequence 20, Appli
41	85.5	8.0	584	10	US-09-815-242-5716 Sequence 5716, Ap
42	85.5	8.0	584	10	US-09-815-242-12476 Sequence 12476, A
43	85.5	8.0	697	10	US-09-881-752A-76 Sequence 76, Appli
44	85.5	8.0	1638	12	US-10-090-458-2 Sequence 2, Appli
45	84	7.8	391	10	US-09-815-242-11566 Sequence 11566, A

ALIGNMENTS

RESULT 1
US-09-815-242-13329
; Sequence 13329, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 13329
; LENGTH: 213
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-815-242-13329

Query Match 100.0%; Score 1071; DB 10; Length 213;
Best Local Similarity 100.0%; Pred. No. 2.8e-95;


```

Qy 1 MITVLLILAYLIGSPSGMWIGOVFFQINLREHSGSNTGTTNTFRILCKKAGMATEVID 60
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 MLIALLLIAYLIGSPGLIVGKLAKGIDIREHSGSNLGNATNARTLGVKAGSVVIAGD 60

Qy 61 FFKGTATLPLIPIHLOGVSPLIFGLLAVIHTGHTFPIAGFKGCKAVATSAGVIFGPAPIF 120
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 61 ILKGTLATAPFLMHVD-IHPLLAGVFAVLGHVFPIFANFKGCKAVATSGVLLFFAPLL 119

Qy 121 CLYLAIITFGALYLGSMISLSSVSTASTAAVIGVLLPFLGFLISNVDSLFIATILLASL 180
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 120 FITWVAFFFLVLTKEVSLSSMLTGITYVI-----YSFFV--HDTVLLIIVVTLTIF 170

Qy 181 IIRHKONIAIRINKNTENLVPM 202

Db 171 VYVHRANIKRIINKTEPKVKW 192

```

```

RESULT 5
US-09-823-246-2
: Sequence 2, Application US/09823246
: Patent No. US20020058789A1
: GENERAL INFORMATION:
: APPLICANT: Burnham, Martin K. R.
: APPLICANT: Biswas, Sanjoy
: APPLICANT: Zalacain, Magdalena
: APPLICANT: Warren, Patrick V.
: APPLICANT: Slyvester, Daniel R.
: APPLICANT: McDevitt, Damien
: TITLE OF INVENTION: ynes
: FILE REFERENCE: GR20001
: CURRENT APPLICATION NUMBER: US/09/823,246
: CURRENT FILING DATE: 2001-03-30
: PRIOR APPLICATION NUMBER: US 60/193,496
: PRIOR FILING DATE: 2000-03-31
: NUMBER OF SEQ ID NOS: 2
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 2
: LENGTH: 202
: TYPE: PRT
: ORGANISM: Staphylococcus aureus
US-09-823-246-2

```

```

Query Match          41.4%; Score 443.5; DB 10; Length 202;
Best Local Similarity 42.9%; Pred. No. 2e-35;
Matches 90; Conservative 43; Mismatches 60; Indels 17; Gaps 3;

QY      1 MITTVLLIAYLIGSPGLWIGOVFFOINLRHSGNGTGTNTFRILGKKAGMATFVID 60
       :|::||::||::||::||::||::||::||::||::||::||::||::||::||
Db      1 MMIIIVMLLSYLICAFPSGVIGKLFKKDKIRQFGSNGTGATSNFVLGRPAGFLVFLFD 60

QY      61 FFKGTALTLPIIPIHLQGVSF-----LIFGLLAVITGHFTFIAPGPKGKAVATSAGV 112
       ||::||::||::||::||::||::||::||::||::||::||::||::||
Db      61 IFKGFTIVFPPLVQLQHADGPSISTFTNGLVIGFLAILGHVPYVLYLKPGQKAVATSAGV 120

QY      113 IFGPAPIFCFLVALTIFFGALVLSMSISSVYTASIAAIVGVLPLFCGLFINNSDLSLIA 172
       :|::||::||::||::||::||::||::||::||::||::||::||::||
Db      121 VLGVNPILLILIATIFFVLKIFKYVSLASIVAACCVISL-----LIQDIYLLVV 173

QY      173 IILASLIIRHKDNTRIARKNTENLVPM 202
       :::::||::||::||::||::||::||::||::||::||::||::||::||
Db      174 FLVSI--ILIIHRHSNTARIERGEPEKIKW 201

```

RESULT 6
US - 09-815-242-5305
: Sequence 5305, Application US/09815242
: Patent No. US2000061569A1
: GENERAL INFORMATION:
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlssen, Kari L.
: APPLICANT: Zyskind, Judith W.
: APPLICANT: Wall, Daniel

```

; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5305
; LENGTH: 202
; TYPE: PR1
; ORGANISM: Staphylococcus aureus
US-09-815-242-5305

```

[illegible]

```

RESULT 7
US - 09-815-242-12137
: Sequence 12137, Application US/09815242
: Patent No. US20020061569A1
: GENERAL INFORMATION:
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlsen, Karl L.
: APPLICANT: Zyskind, Judith W.
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John D.
: APPLICANT: Carr, Grant J.
: APPLICANT: Yamamoto, Robert T.
: APPLICANT: Xu, H. Howard
: TITLE OF INVENTION: Identification of
: TITLE OF INVENTION: Prokaryotes
: FILE REFERENCE: ELITRA.011A
: CURRENT APPLICATION NUMBER: US/09/815,
: CURRENT FILING DATE: 2001-03-21
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848

```


Db 8 MILIAYLCGSISSAILVCRICGLPDPRTSGSGNPGATNVLRIKCGKAVALIFDVILKG- 66
Qy 66 LATLLPI--IFHLOGVSPILFGLLAV---IGHTFPPIFAGKGGKAVATSGVIFGPAIF 120
Db 67 ---MLPVNGAYEL-GVSPFWLGLIAIAACLGHIWPFVFGKGGKGVATAGAI---API- 118
Qy 121 CLYLAIFFGALYLGSMISLSSVASTAAVIGVLLFPLFGFILSNYDSLFIAIILASL 180
Db 119 GWDLTGVWAGTWLTVILSGYS---SLGAIUSALIAP---FYVMFKPQTFPPVSMUSCL 172
Qy 181 IIRHKDNIARIKKNKTENLVPW 202
Db 173 ILLRHDNIORLWRROETKI-W 193

RESULT 10

US-09-815-242-11713
; Sequence 11713, Application US/09815242
; Patent No. US20020061569A1

GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in

FILE REFERENCE: ELITRA 011A

CURRENT APPLICATION NUMBER: US/09/815,242

CURRENT FILING DATE: 2001-03-21

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

NUMBER OF SEQ ID NOS: 14110

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 11713

LENGTH: 205

TYPE: PRT

ORGANISM: Klebsiella pneumoniae

US-09-815-242-11713

Query Match 23.6%; Score 252.5; DB 10; Length 205;
Best Local Similarity 35.6%; Pred. No. 3.4e-17;
Matches 72; Conservative 40; Mismatches 69; Indels 21; Gaps 9;

Qy 6 LLIILAYLIGSPISGLWIGOVFFQINLREHSGNTGTTNTRFILCKKAGMATFFVDFKGT 65
Db 8 LVLIAYLCGSISSAILVCRICGLPDPRTSGSGNPGATNVLRIKCGKAVALIFDVILKG- 66
Qy 66 LATLLPI--IFHLOGVSPILFGLLAV---IGHTFPPIFAGKGGKAVATSGVIFGPAIF 120
Db 67 ---MLPVNGAWAL-GLTFPFWLGLVAIAACVGIWPFVFFHFGKGGKGVATAGAI---API- 118
Qy 121 CLYLAIFFGALYLGSMISLSSVASTAAVIGVLLFPLFGFILSNYDSLFIAIILASL 180
Db 119 GLDITGVWAGTWLTVILSGYS---SLGAIUSALIAP---FYVMFKPQTFPPVSMUSCL 172
Qy 181 IIRHKDNIARIKKNKTENLVPW 202

Db 173 ILLRHDNIORLWRROESKI-W 193

RESULT 11

US-09-815-242-10997
; Sequence 10997, Application US/09815242
; Patent No. US20020061569A1

GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in

FILE REFERENCE: ELITRA 011A

CURRENT APPLICATION NUMBER: US/09/815,242

CURRENT FILING DATE: 2001-03-21

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

NUMBER OF SEQ ID NOS: 14110

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 10997

LENGTH: 199

TYPE: PRT

ORGANISM: Haemophilus influenzae

US-09-815-242-10997

Query Match 22.8%; Score 244.5; DB 10; Length 199;
Best Local Similarity 33.0%; Pred. No. 1.9e-16;
Matches 68; Conservative 41; Mismatches 78; Indels 19; Gaps 7;

Qy 1 MITIVLLIAYLALGSPISGLWIGOVFFQINLREHSGNTGTTNTRFILCKKAGMATFFVDF 60
Db 3 LFAIFYMLFAYLIGSISAILICRIAGLPDPRTSGSGNPGATNVLRIKGNKKSALAVLIFD 62
Qy 61 FFKGTATLLPI-IFHLOGVSPILFGLLAV---IGHTFPPIFAGKGGKAVATSGVIFGF 116
Db 63 MLKG---MIPVWAGYVLGLTQFELGNVALGACLGHIFFIPIFFKGGKGVATAGAI--- 115

Qy 117 APIFCYLAIIFFGALYLGSMISLSSVASTAAVIGVLLFPLFGFILSNYDSLFIAIILA 176
Db 116 APISWAVAGSMFGTWIF---VELVSGYSLSLAVSALLVPFYVMFKPQTFPPVALV-- 169
Qy 177 LASLIIRHKDNIARIKKNKTENLVPW 202
Db 170 -CCLLIYRHDNIORLWRGOEDKV-W 193

RESULT 12

US-09-815-242-11799
; Sequence 11799, Application US/09815242
; Patent No. US20020061569A1

GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.

APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT FILING DATE: 2001-03-21
CURRENT APPLICATION NUMBER: US/09/815,242
PRIORITY FILING DATE: 2001-03-21
PRIORITY APPLICATION NUMBER: 60/191,078
PRIORITY FILING DATE: 2000-03-21
PRIORITY APPLICATION NUMBER: 60/206,848
PRIORITY FILING DATE: 2000-05-23
PRIORITY APPLICATION NUMBER: 60/207,727
PRIORITY FILING DATE: 2000-05-26
PRIORITY APPLICATION NUMBER: 60/242,578
PRIORITY FILING DATE: 2000-10-23
PRIORITY APPLICATION NUMBER: 60/253,625
PRIORITY FILING DATE: 2000-11-27
PRIORITY APPLICATION NUMBER: 60/257,931
PRIORITY FILING DATE: 2000-12-22
PRIORITY APPLICATION NUMBER: 60/269,308
PRIORITY FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 11799

Query Match 22.3%; Score 239; DB 10; Length 189;
Best Local Similarity 34.2%; Pred. No. 6.1e-16;
Matches 65; Conservative 28; Mismatches 77; Indels 20; Gaps 3;

Y 5 VLLILAYLGSIPSGMWGVFQINLRHSGSGNTGTTNFRILGKAGKAGMATVDFK 64
D 4 LLAILAYLGSLSFVALLSFWFTQDPRASGSGNPGATNMLRVAGKLAITLLGVDGK 63
Y 65 TLATLPIIFHAGVSLPIFLGLAVIGHTFPFAGFKGKAVATSAGVIFGAPICLYL 124
D 64 LLPVLVARGLGVMEENWYGAIVGHLYPLFNFPGKGAVATAAGMLGLYPPAVLLA 123
Y 125 AIIFFGALYLGSMISLSSVTSIAAIVGVLLPFLF-----GFILSNYDSLFIAIILAS 179
D 124 AAALLTFKLSRTSSLASLAVT-----PLTLLAWQOPGALLP-----MTVLTG 168
Y 180 LIIRHKDNI 189
D 169 LIVVHRANL 178

RESULT 13

US-09-815-242-11614
Sequence 11614, Application US/09815242
Patent No. US20020061569A1

GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in
TITLE OF INVENTION: Prokaryotes

FILE REFERENCE: ELITRA.011A

CURRENT APPLICATION NUMBER: US/09/815,242

CURRENT FILING DATE: 2001-03-21

PRIORITY APPLICATION NUMBER: 60/191,078

PRIORITY FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 11614

LENGTH: 220

TYPE: PRT

ORGANISM: Helicobacter pylori

US-09-815-242-11614

Query Match 19.9%; Score 213; DB 10; Length 220;

Best Local Similarity 26.6%; Pred. No. 2.2e-13;

Matches 58; Conservative 45; Mismatches 87; Indels 28; Gaps 6;

Y 2 ITIVLLILAYLGSIPSGMWGVFQINLRHSGSGNTGTTNFRIL-----GKAGM 54

D 11 INVIFTLGLYIGGIPFGYALMKIFYGMDITKIGSGGIGATNVLRLQSKGVSNKQMAL 70

Y 55 ATFTVIDFFKGTATLPIIFHAGVSLPIFLGLAVIGHTFPFAGFKGKAVATSAGVIF 114

D 71 LVLLDLFKGMFAVFLSKLFGDLYSLQWVAIASILGHCHYSPFLNFGNGKGVSTINGSVV 130

Y 115 GFAPIFCLYLAIFFGALYLGSMISLSSVTSIAAIVG-----VLLF-----PLFGFI 162

D 131 LLPIESLIGLTVWF---FVGKVLKIS---SLASILGVGTATVLVIFFPYMHIPDSVNI 183

Y 163 LSNYDSLFTAILALASLIIIRHKONIAIKKNNTENLV 200

D 184 LREVGOTPMVLIFITFL--INKHAGNIFNLLTGCKKKV 219

RESULT 14

US-09-815-242-11453

Sequence 11453, Application US/09815242

Patent No. US20020061569A1

GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in
TITLE OF INVENTION: Prokaryotes

FILE REFERENCE: ELITRA.011A

CURRENT APPLICATION NUMBER: US/09/815,242

CURRENT FILING DATE: 2001-03-21

PRIORITY APPLICATION NUMBER: 60/191,078

PRIORITY FILING DATE: 2000-03-21

PRIORITY APPLICATION NUMBER: 60/206,848

PRIORITY FILING DATE: 2000-05-23

PRIORITY APPLICATION NUMBER: 60/207,727

PRIORITY FILING DATE: 2000-05-26

PRIORITY APPLICATION NUMBER: 60/242,578

PRIORITY FILING DATE: 2000-10-23

PRIORITY APPLICATION NUMBER: 60/253,625

PRIORITY FILING DATE: 2000-11-27

PRIORITY APPLICATION NUMBER: 60/257,931

PRIORITY FILING DATE: 2000-12-22

PRIORITY APPLICATION NUMBER: 60/269,308

PRIORITY FILING DATE: 2001-02-16

NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 11453
LENGTH: 262
TYPE: PRT
ORGANISM: Helicobacter pylori
US-09-815-242-11453

Query Match 19.7%; Score 211; DB 10; Length 262;
Best Local Similarity 27.1%; Pred. No. 4.2e-13;
Matches 56; Conservative 44; Mismatches 79; Indels 28; Gaps 6;
QY 2 ITIVLILAYLIGSIPSGLWIGVFOINLREHSGSGNTGTTNTRIL-----GKKAGM 54
DB 53 INVITFLGGLGGIPFGYALMKIFGYMDITKIGSGGIGATNVLALOSKGVSNKQMAL 112
QY 55 ATFVIDFFKGTATLLPIIFHLOGVSPILFGLLAVIGHTFPFAGFKGKAVATSAGVIF 114
DB 113 LVLILDLPKMEAVFLSKFLGLDYSLOWMVAIASILGHCSPLNPNFGKGVSTIMGSV 172
QY 115 GFAPICFLYLAIFGALYLSMISLSSVTSIAAIVG-----VLLF-----PLFGFI 162
DB 173 LLIPIESLGLTWTF---FVGKVLKIS-----SLASILGVGTATVLIFFPYMHIPDSVNI 225
QY 163 LSNYDSLFIATIALASLIIRHKONI 189
DB 226 LAEVGTQTPWVLIFITFL--IKHAGNI 250

RESULT 15
US-09-778-927A-79
Sequence 79, Application US/09778927A
Patent No. US2002008342A1
GENERAL INFORMATION:
APPLICANT: KHOSRAVI, Rami et al.
TITLE OF INVENTION: NOVEL NUCLEIC ACID AND AMINO ACID SEQUENCES AND NOVEL
VARIANTS OF ALTERNATIVE SPLICING
FILE REFERENCE: 2786-0160P
CURRENT APPLICATION NUMBER: US/09/778, 927A
CURRENT FILING DATE: 2001-02-08
PRIOR APPLICATION NUMBER: IL 134453
PRIOR FILING DATE: 2000-02-09
PRIOR APPLICATION NUMBER: IL135341
PRIOR FILING DATE: 2000-03-29
NUMBER OF SEQ ID NOS: 81
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 79
LENGTH: 434
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(434)
OTHER INFORMATION: xaa = any amino acid, unknown, or other
US-09-778-927A-79

Query Match 9.9%; Score 105.5; DB 10; Length 434;
Best Local Similarity 25.0%; Pred. No. 0.0088;
Matches 53; Conservative 33; Mismatches 73; Indels 53; Gaps 12;
QY 11 YLIGSIPSGL-----WIG--OVFOINLREHSGSGNTGTTNTRILG-----KKAGMA 55
DB 5 WLLGDDPAGSAGTGPWLRSREQVFLISA---ASVNLGSMCMYSLGPPFPKAEKKGAS 61
QY 56 TFVIDFFKGTATLLPIIFHLOGVSPILFGLLAVIGHTFPFAGFKGKAVATSAGVIF 114
DB 62 NTIIGMIFGCA-----LFEL--LASLVFNGVLVHIGAKFM---FVAGMEVSGGVITLF 110
QY 115 GF-----APIF---CLYL-----AIIFFGALYLSMISLSSVTSIAAIVG-----152
DB 111 GVLDRVPDGPVFITAMFLVRVMDAVSFAAAMTASSILAKAFPNPNVATVLGSLETFSLG 170
QY 153 -VLLFPFLGFIILSNYDSLFIATIALASLIIRHKONI 189

Db 171 LILGPPVGGFLYQSGVEVPFIVLGCVVLLMV 202
Search completed: December 26, 2002, 02:58:43
Job time : 90.1872 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 25, 2002, 14:27:20 : Search time 151.618 Seconds
(without alignments)
58.268 Million cell updates/sec

Title: US-10-068-080-1
Perfect score: 1071
Sequence: 1 MITIVLLILAVLGSIPSL.....NKTENLVPWGLNLTHODPKK 213

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	1071	100.0	213	1	Y851_STRPN	Q54916	streptococ
2	468	43.7	193	1	YNES_BACSU	Q45064	bacillus su
3	332	31.0	192	1	Y676_AQUAE	Q66905	aquifex aeo
4	300.5	28.1	222	1	YJ73_SYNV3	P73933	synecocyst
5	263.5	24.6	205	1	YGIH_ECOLI	P31056	escherichia
6	244.5	22.8	199	1	YGIH_HAEIN	P44603	haemophilus
7	223.5	20.9	239	1	Y247_MYCCE	P47489	mycoplasma
8	214.5	20.0	239	1	Y247_MYCPN	P75428	mycoplasma
9	213	19.9	220	1	YF09_HELPJ	Q921b1	helicobacte
10	211	19.7	220	1	YF09_HELPJ	Q921b1	helicobacte
11	117	10.9	304	1	YF09_HELPJ	Q921b1	helicobacte
12	102	9.5	348	1	YF09_HELPJ	Q921b1	helicobacte
13	97.5	9.1	330	1	Y374_METJA	Q78680	carassius a
14	95.5	8.9	173	1	Y374_METJA	Q78680	carassius a
15	95.5	8.9	173	1	Y374_METJA	Q78680	carassius a
16	95.5	8.9	173	1	Y374_METJA	Q78680	carassius a
17	95.5	8.9	173	1	Y374_METJA	Q78680	carassius a
18	95.5	8.9	173	1	Y374_METJA	Q78680	carassius a
19	94.5	8.8	414	1	Y0AB_BACSU	Q34864	bacillus su
20	93	8.7	598	1	CSTA_BACSU	P44532	bacillus su
21	92.5	8.6	458	1	YUON_RICCN	Q92hh5	ricketsia
22	92.5	8.6	492	1	YD19_METJA	Q58715	methanococ
23	91.5	8.5	412	1	Y663_METH	Q28759	methanobact
24	91.5	8.5	438	1	PBOX_BACSU	P42086	bacillus su
25	91.5	8.5	552	1	CYHR_CANMA	P32071	candida mal
26	91	8.5	497	1	DTPT_LACLA	Q9mix9	brachydanio
27	90.5	8.5	172	1	NU6M_BRARE	P36574	lactococcus
28	90.5	8.5	220	1	YQJA_ECOLI	P42614	escherichia
29	90.5	8.5	442	1	CITN_LACLA	P21608	lactococcus
30	90.5	8.5	452	1	YDUE_ECOLI	P38055	escherichia
31	90	8.4	276	1	NO5Y_PSEST	P19845	pseudomonas
32	89.5	8.4	423	1	YHJV_ECOLI	P37660	escherichia
33	89.5	8.4	440	1	CAPE_STAAU	P39854	staphylococ

34	89	8.3	201	1	PSS_METJA	Q58609	methanococ
35	89	8.3	473	1	MOT4_CHICK	P57788	gallus gall
36	89	8.3	622	1	COX1_BACSU	P24010	bacillus su
37	88.5	8.3	170	1	NUOJ_BUCAI	P57260	buchnera ap
38	88.5	8.3	428	1	DTA_ECOLI	Q8x5m2	escherichia
39	88.5	8.3	428	1	DTA_ECOLI	P37312	escherichia
40	88.5	8.3	439	1	IDNT_ECOLI	P39344	escherichia
41	88.5	8.3	546	1	YAO5_SCHPO	Q10084	schizosacch
42	88	8.2	162	1	Y159_PYRHO	O57898	pyrococcus
43	88	8.2	413	1	YGL8_AQUAE	O67545	aquifex aeo
44	88	8.2	450	1	GNUT_PSEAE	Q921j1	pseudomonas
45	88	8.2	450	1	VGLM_HSVB	P28948	equine herp

ALIGNMENTS

RESULT 1
Y851_STRPN
ID Y851_STRPN STANDARD; PRT; 213 AA.
AC Q54916;
DT 15-JUL-1999 (Rel. 38, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein SP0851.
GN SP0851.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=7785;
RX MEDLINE=96313246; PubMed=8763932;
RA Pan X., Fisher M.;
RT "Cloning and characterization of the parC and parE genes of
Streptococcus pneumoniae encoding DNA topoisomerase IV: Role in
fluoroquinolone resistance.";
RT J. Bacteriol. 178:4060-4069(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tetelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
pneumoniae.";
RT Science 293:498-506(2001).
CC -!- SUPCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE UPF0078 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z67739; CAA91549.1; -
CC EMBL; AF007390; AAK74980.1; -
CC TIGR; SP0851; -
CC InterPro; IPR003811; DUF205.
CC Pfam; PF02660; DUF205; 1.
CC TIGRFAMs; TIGR00023; DUF205; 1.
CC Hypothetical protein; Transmembrane; Complete proteome.
CC TRANSMEM 2 22 POTENTIAL.

```
FT TRANSMEM 70 90 POTENTIAL.
FT TRANSMEM 112 132 POTENTIAL.
FT TRANSMEM 143 163 POTENTIAL.
FT TRANSMEM 164 183 POTENTIAL.
FT CONFLICT 168 184 S -> F (IN REF. 1).
SQ SEQUENCE 213 AA; 22928 MW; 22CB089C17750818 CRC64;

Query Match 100.08; Score 1071; DB 1; Length 213;
Best Local Similarity 100.08; Pred. No. 3.4e-74;
Matches 213; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MITVILLIAYLGSTPSGLWIGQVFFQINLREHSGNGTGTNTFRILKKGAGMATFVID 60
    |||||
Db 1 MITVILLIAYLGSTPSGLWIGQVFFQINLREHSGNGTGTNTFRILKKGAGMATFVID 60
    |||||

Qy 61 FFKGTLATLLPIIFHLQGVSPFLIFGLAVIGHTFTPIFAGKGGKAVATSGVIFGAPIF 120
    |||||
Db 61 FFKGTLATLLPIIFHLQGVSPFLIFGLAVIGHTFTPIFAGKGGKAVATSGVIFGAPIF 120
    |||||

Qy 121 CLYLAIIFGALYLGSMISLSSVTASIAAVICVLLFPLFGFILTNSYDSLFTAILALASL 180
    |||||
Db 121 CLYLAIIFGALYLGSMISLSSVTASIAAVICVLLFPLFGFILTNSYDSLFTAILALASL 180
    |||||

Qy 181 IIIRHKDNIARIKKNKTENLVPWGLNTHODPKK 213
    |||||
Db 181 IIIRHKDNIARIKKNKTENLVPWGLNTHODPKK 213
    |||||

RESULT 2
YNES_BACSU STANDARD; PRT: 193 AA.
AC Q45064;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein yneS.
GN YNES.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RC STRAIN=168;
RX MEDLINE=97124194; PubMed=8969507;
RA Rose M., Entian K.D.;
RT "New genes in the 170 degree region of the Bacillus subtilis genome
RT encode DNA gyrase subunits, a thioredoxin, a xylanase and an amino
RT acid transporter."
RL Microbiology 142:3097-3101(1996).
RN [2]
RC STRAIN=FROM N.A.
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Boriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Chou S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasehara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medona N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,
RA Neelne D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Purnelle S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
```

```
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Zumbstein A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256(1997).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE UPF0078 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Z73234; CAA97604.1;
CC EMBL: Z99113; CAB13690.1;
CC Subtilist; BGI1826; yneS.
CC InterPro: IPR003811; DUF205.
CC Pfam: PF02660; DUF205; 1.
CC TIGRFAMs: TIGR00023; DUF205; 1.
CC KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 2 22 POTENTIAL.
FT TRANSMEM 54 74 POTENTIAL.
FT TRANSMEM 76 96 POTENTIAL.
FT TRANSMEM 111 131 POTENTIAL.
FT TRANSMEM 153 173 POTENTIAL.
SQ SEQUENCE 193 AA; 20966 MW; C75803C399B97292 CRC64;

Query Match 43.78; Score 468; DB 1; Length 193;
Best Local Similarity 48.08; Pred. No. 7.9e-29;
Matches 97; Conservative 32; Mismatches 63; Indels 10; Gaps 3;

Qy 1 MITVILLIAYLGSTPSGLWIGQVFFQINLREHSGNGTGTNTFRILKKGAGMATFVID 60
    |||||
Db 1 MLJALLIAYLGSTPSGLWIGQVFFQINLREHSGNGTGTNTFRILKKGAGMATFVID 60
    |||||

Qy 61 FFKGTLATLLPIIFHLQGVSPFLIFGLAVIGHTFTPIFAGKGGKAVATSGVIFGAPIF 120
    |||||
Db 61 ILKGTALATLPLMHVD-IHPLLACVAVLGHVFPFAKFKGKAVATSGVLLFYAPLL 119
    |||||

Qy 121 CLYLAIIFGALYLGSMISLSSVTASIAAVICVLLFPLFGFILTNSYDSLFTAILALASL 180
    |||||
Db 120 FITMVAVFFIFLYLTKFVLSLSMLTGIVTI-----YSFFV--HDTYLLIVVTLTIF 170
    |||||

Qy 181 IIIRHKDNIARIKKNKTENLVPW 202
    |||||
Db 171 VIYHRANKRIINKTEPKVKW 192
    |||||

RESULT 3
Y676_AQUAE STANDARD; PRT: 192 AA.
AC O66905;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein AQ_676.
GN AQ_676.
OS Aquifex aeolicus.
OC Bacteria; Aquificae; Aquificae (class); Aquificales; Aquificaceae;
OC Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
```

```

RA MEDLINE=98196666; PubMed=9537320;
RA Decker G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Aujeay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus.";
RL Nature 392:353-358(1998).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE UPF0078 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE000701; AAC06869.1; -
DR InterPro: IPR003811; DUF205.
DR Pfam: PF02660; DUF205. 1.
DR TIGRFAMs: TIGR00023; DUF205; 1.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 3 23 POTENTIAL.
FT TRANSMEM 51 71 POTENTIAL.
FT TRANSMEM 80 100 POTENTIAL.
FT TRANSMEM 112 132 POTENTIAL.
FT TRANSMEM 149 169 POTENTIAL.
SQ SEQUENCE 192 AA; 20940 MW; EAD53C4016D63C00 CRC64;

Query Match 31.0%; Score 332; DB 1; Length 192;
Best Local Similarity 39.8%; Pred. No. 1.4e-18;
Matches 80; Conservative 39; Mismatches 68; Indels 14; Gaps 6;

QY 1 MITIVLLILAYLLGSLWIGVFFQINLREHSGNLTGTNTFRILGKKGAGMATFVID 60
DB 1 MKALFLVIFAYLLGSITTFGEVIKLV-KGVDLNRVSGNMGVATNTVTRALGKGVGLVLFELD 59

QY 61 FPKGTALTLPI-IFHLOQVSLPIFGLAVIGHFTPIFAGFGKGVATSGVIFGFAP 119
DB 60 FUKGIPALIAVKSFGIDSWLTFGLASVLGHMYPVFVFGKGGKGVATGVLGVFAVSPS 119

QY 120 FCLYLAIIFFGALYLGSMISLSSVTASIAAVIGVLLFPLFGFILSNY--DSLFIATILAL 177
DB 120 VALFSFLWVLGIFLWKRVSLASITATISAF-----LFLFV-AGYVPNVLFMAIV--I 169

QY 178 ASLIIIRHKDNIARTKNNTN 198
DB 170 GALIIRHRENINRLTGREH 190

RESULT 4
YJ73_SYNY3
ID AC YJ73_SYNY3 STANDARD; PRT; 222 AA.
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein sll1973.
GN SLL1973.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirosewa M., Sugiyama M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the

```

```

RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE UPF0078 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D90910; BAA17999.1; -
DR InterPro: IPR003811; DUF205.
DR Pfam: PF02660; DUF205. 1.
DR TIGRFAMs: TIGR00023; DUF205; 1.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 4 24 POTENTIAL.
FT TRANSMEM 56 76 POTENTIAL.
FT TRANSMEM 87 107 POTENTIAL.
FT TRANSMEM 130 150 POTENTIAL.
FT TRANSMEM 153 173 POTENTIAL.
SQ SEQUENCE 222 AA; 23534 MW; 4718AB2C20833360 CRC64;

Query Match 28.1%; Score 300.5; DB 1; Length 222;
Best Local Similarity 34.2%; Pred. No. 3.7e-16;
Matches 75; Conservative 37; Mismatches 66; Indels 41; Gaps 6;

QY 1 MITIVLLILAYLLGSLWIGVFFQINLREHSGNLTGTNTFRILGKKGAGMATFVID 60
DB 6 LLCICLLIITLMSIPGYLAGLLGIDIREHSGKSTGATNVRTLCKPAAIAVLAD 65

QY 61 FFKGTALTLPIIFHLOQVSLPI-----FGLLAVIGHFTPIFAGFGKGVATSG 109
DB 66 ISKGVMAVALVRAIYSGDMLPALPAQWNLTLGVAVIAVLGHKSKSIFLKFSGGKSVATS 125

QY 110 AGVIFGFAPICFLYLAI-----IFFGALYLGSMISLSSVTASIAAVIGVLL 158
DB 126 LGVLV-----MNLIMWALGTALFTLVIFTRIVSLSSVAAI-AVNGIALALQLPPPYLA 180

QY 159 FGFILNSDLSFIATILALASLIIRHKDNIARTKNKTE 197
DB 161 FTFLAGMY-----VIVRHTNIERTILOGTE 205

RESULT 5
YGIH_ECOLI
ID YGIH_ECOLI STANDARD; PRT; 205 AA.
AC P31056;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein ygiH.
GN YGIH OR B3059 OR 24412 OR ECS3942.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RN Science 277:1453-1474(1997).
RP [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;

```



```
QY 61 FFKGTLATLPI-IPHLOGVSLIFGLAV---IGHTEPIFAGKGGKAVATAGVIFGF 116
Db 63 MKUG-----MIPWAGYIIGLTOFELGVALGACLGHIPIFOFAGGAGVATAGAI--- 115
QY 117 APIFLYLAIIFFGALYLGSMISLSSVTASIAAVIGVLFFLFGFILSNYDSLFIATILA 176
Db 116 APISNAVAGSMFGTWIF-----VFLVSGYSSLSAVISALLVPFYVMWPKPEFTEPVALV-- 169
QY 177 LASLIIRHKONIAIRIKNKTENLVPW 202
Db 170 -CCLLIYRHNDIORLWRGOEDKV-W 193

RESULT 7
Y247_MYCGE STANDARD; PRT; 239 AA.
AC P47489;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MG247.
GN MG247.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA Tomb J.-F., Ufferback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.-F., Dougherty B.A., Bost K.F., Hu P.-C., Lucier T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "The minimal gene complement of Mycoplasma genitalium.";
RL Science 270:397-403(1995).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE UPF0078 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U39703; AAC71467.1;
DR TIGR: MG247;
DR InterPro: IPR003811; DUF205.
DR Pfam: PF02660; DUF205; 1.
DR TIGRFAMs: TIGR00023; DUF205; 1.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 7 27 POTENTIAL.
FT TRANSMEM 61 81 POTENTIAL.
FT TRANSMEM 99 119 POTENTIAL.
FT TRANSMEM 135 155 POTENTIAL.
FT TRANSMEM 159 179 POTENTIAL.
FT TRANSMEM 199 219 POTENTIAL.
SQ SEQUENCE 239 AA; 27489 MW; D78CE976DEF621FD CRC64;

Query Match 20.9%; Score 223.5; DB 1; Length 239;
Best Local Similarity 31.8%; Pred. No. 2.5e-10;
Matches 74; Conservative 33; Mismatches 87; Indels 39; Gaps 10;

QY 1 MITIVLLIAYLLGSPSGLMIGVFFOI---NLREHSGNGTGTNTFRILCKKAGMATF 57
Db 9 ILVIFSLASGYLGST-----IFADIFSKILKNVREFGSKNPGATNSRVFLGKIGFLVA 64
QY 58 VIDFFKGLTATLPIIT---FHLQGV-----SPLIFGLL-----AVIGHTFFIFAGFKGG 103
Db 65 IFDAFKGFFAFLITWLIFRFGLOGYLTERKVGQSTFVLSLSCFAATIGHIPFLYKFKGG 124

QY 61 FFKGTLATLPI-IPHLOGVSLIFGLAV---IGHTEPIFAGKGGKAVATAGVIFGF 116
Db 63 MKUG-----MIPWAGYIIGLTOFELGVALGACLGHIPIFOFAGGAGVATAGAI--- 115
QY 117 APIFLYLAIIFFGALYLGSMISLSSVTASIAAVIGVLFFLFGFILSNYDSLFIATILA 176
Db 116 APISNAVAGSMFGTWIF-----VFLVSGYSSLSAVISALLVPFYVMWPKPEFTEPVALV-- 169
QY 177 LASLIIRHKONIAIRIKNKTENLVPW 202
Db 170 -CCLLIYRHNDIORLWRGOEDKV-W 193
```

```
QY 104 KAVATSAGVIFGPAIFCLYLAIIFFGALYLGSMISLSS-VTASIAAVIGVLFFLPGFI 162
Db 125 KAIATGGSLAISLWMLICLLIWMITLITKYVSLASLITFFVLAVI--ILIPWLDYL 182
QY 163 -LSNYDSL-----FIATILA-----LASLIIRHKONIAIRIKNKTENLV 200
Db 183 YFFNSDPLKSYQNEWYIILFFCLWYWPVTVVVFWLHRANIIRLHCKESKI 235

RESULT 8
Y247_MYCPN STANDARD; PRT; 239 AA.
AC P75428;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MG247 homolog (H91_orf239).
GN MPN350 OR MP486.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelreich R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,
RA Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae.";
RL Nucleic Acids Res. 24:4420-4449(1996).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE UPF0078 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE000047; AAB96134.1;
DR InterPro: IPR003811; DUF205.
DR Pfam: PF02660; DUF205; 1.
DR TIGRFAMs: TIGR00023; DUF205; 1.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 6 26 POTENTIAL.
FT TRANSMEM 62 82 POTENTIAL.
FT TRANSMEM 99 119 POTENTIAL.
FT TRANSMEM 135 155 POTENTIAL.
FT TRANSMEM 159 179 POTENTIAL.
FT TRANSMEM 199 219 POTENTIAL.
SQ SEQUENCE 239 AA; 27439 MW; 6D4110A8253C9EBB CRC64;

Query Match 20.0%; Score 214.5; DB 1; Length 239;
Best Local Similarity 29.2%; Pred. No. 1.2e-09;
Matches 69; Conservative 40; Mismatches 82; Indels 45; Gaps 10;

QY 1 MITIVLLIAYLLGSPSGLMIGVFFOI---NLREHSGNGTGTNTFRILCKKAGMATF 57
Db 9 LLIVSLVIGILMGSV---MFADYFGKILNKDVRKLSKNPGATNSIRVFLGKIGFLVG 64
QY 58 VIDFFKGLTIA---TLLPIIFHLQGVSL-----IFGLLAVIGHTFFIFAGFKGG 103
Db 65 LCDALKGLFLAFVSEFLSFMLQYLLNVNQYKVYLYTLVSLSCFAATIGHIPFLYKFKGG 124
QY 104 KAVATSAGVIFGPA---IFCLYLAIIFFGALYLGSMISLSS-VTASIAAVIGVLFFLPLF 159
Db 125 KAIATGGSLAISLWMLFVICVLWLL---VTLITKYVSLASLITFFLITAI--ILVPL 179
QY 160 GPI-----LSNYDSLFIATILA-----LASLIIRHKONIAIRIKNKTENLV 200
```



```

NCBI_TaxID=2190;
[1]
SEQUENCE FROM N.A.
RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.-J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.F., Merrick J.M., Glodek A.,
RA Scott J.L., Kirchhausen N.S.M., Weidman J.G., Fuhrmann J.L., Nguyen D.,
RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst H.O., Kane B.P., Borodovsky M.,
RA Klenn H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT *Complete genome sequence of the methanogenic archaeon, Methanococcus
RL jannaschii;.
RR Science 273:1058-1073(1996).
CC -|- SIMILARITY: TO M_JANNASCHII MJ0201 AND TO H_INFLUENZAE AND
CC H_SONNUS HI0703.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to licenselsb@sib.ch).
CC -----
DR EMBL: U67490; AACB9363.1; .
DR TRIGR: MJ0374; .
DR InterPro: IPR000252; DedA.
DR InterPro: IPR000326; PA_PTPase.
DR Pfam: PF00597; DedA; 1.
DR Pfam: PF01569; PAP2; 1.
DR SMART: SM00014; acidPPC; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 330 AA; 37018 MW; BA84B0170694C097 CRC64;
Query Match          9.1%; Score 97.5; DB 1; Length 330;
Best Local Similarity 23.0%; Pred.No. 1;
Matches 47; Conservative 29; Mismatches 43; Indels 85; Gaps 9;
OY      57 FVIDPFKGLATLLPIFIHQ-----GVSPLIIFGLAVIGHTF-----PIFA 98
DB       :|||   : |||   : |::|   : |||   : |||
DB     26 FILSFEAFIQIPDPDVFVITAGSFQGLNPITSIAIVATIGTLGGFLGYFLGDKLGHPIFI 85
OY      99 GKPKGRAV-----ATSAGVIFFGP-----IPL-----122
DB       |||    ::|||
DB     86 KLGEKYLNKGEEFNKNYGVGVIAGFSPLPYKVIAWLSGIPEMKHLFTVTGTIGLRLP 145
OY     123 --YLAIFFGALYGMISLVSTASTAAVIGVLFPFLFGFILSNYSDFTAI-----173
DB        :||| : ||| : ||| : ||| : ||| : ||| : |||
DB    146 RLFAVAIFGCV-LGNINRLSDIN-----IYLFYNISHNYIFEADIMPIISKTA 193
OY     174 --ILALASLIIRHKDNRIARKNK 195
DB         :| : ||| : |||
DB    194 YPLIAITSLIIF-----IKNR 209

RESULT 14
NU6M_LATCH ID NU6M_LATCH STANDARD; PRT; 173 AA.
AC AC03175;
DT DT 15-JUL-1998 (Rel. 36, Created)
DT DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE NADH+ubiquinone oxidoreductase chain 6 (EC 1.6.5.3).
GN MTND6 OR ND6 OR NADH6.
OS Latimeria chalumnae (Latimeria) (Coelacanth).
OC Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Coelacanthiformes; Coelacanthidae; Latimeria.
OX NCBI_TaxId=7897;
RN [1]
```

CC or send an email to license@isb-sib.ch).

EMRI: AF006597: AAK34425.1: -

Query Match 8.98: Score 95.5: DB 1: Length 336:

2 ITTVLITLAIYIGSIPSGLWIGOVFFOI-----NLREHSGSNTGTTNTRILGKK 51

Search completed: December 26, 2002. 00:54:14

Search completed: Decem
Job time : 152.618 secs

Result No.	Score	Query Match %	Length	DB	ID	Description
1	716.5	66.9	213	16	O9CGW4	O9cgw4 lactococcus
2	716	65.9	213	15	O9A070	O9a070 streptococc
3	671	58.0	168	2	O9X972	O9x972 streptococc
4	463.5	43.3	198	16	O8Y7J3	O8y7j3 listeria mo
5	459.5	42.9	198	16	O92C68	O92c68 listeria in
6	441.5	41.2	202	16	O99UC5	O99uc5 staphylococ
7	385	35.9	194	16	O8RFY9	O8rfy9 staphylococ
8	311.5	29.1	198	16	O8R9J2	O8r9j2 thermomanae
9	298	27.8	206	16	O9KCD3	O9kcd3 bacillus ha
10	291.5	27.2	198	16	O9RSV1	O9rsv1 deinococcus
11	286.5	26.8	195	16	O9BW84	O9bw84 rhizobium l
12	282.5	26.4	201	16	O8Y664	O8yc64 brucella me
13	282	26.3	205	16	O8URU1	O8uruf agrobacteri
14	275.5	25.7	198	16	O97G69	O97g69 clostridium
15	275.5	25.7	200	16	O9JUL4	O9jul4 nisseria m
16	274.5	25.6	199	2	O9ZAF3	O9zaf3 thermus m

```
Db 61 DLLKCTLATLPLFFHNGVSPILFGLLAVIGHTFISIFDRFKGKAVATSAGVILGFSPL 120
QY 120 FCLYLAIIFGALYGLSISLSSVTASTAAVIGVLLPFLFGFILSNYSDFIAIILALAS 179
Db 121 FLIYLLVVFIIIVLWLFMSISLSSVIGAVFALLGILFISIGFILTSYDLISIIIFVLAI 180
QY 180 LIIIRHKDNIARIKNNKTNLVPWGLNLTHODPKK 212
Db 181 IILRHRTNLRKNNKNCESLVPFGLNLSKQEK 213

RESULT 2
Q9A070 PRELIMINARY: PRT: 213 AA.
ID Q9A070 AC
AC Q9A070:
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein SPY0908.
GN SPY0908.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SF370 / ATCC 700294 / SEROTYPE M1;
RX MEDLINE=21192684; PubMed=11296296;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savić G., Lyon K.,
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
DR EMBL; A006539; AAK33824.1; -;
DR InterPro; IPR003811; DUF205.
DR Pfam; PF02660; DUF205; 1.
DR TIGRFAMs; TIGR00023; DUF205; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 213 AA; 23369 MW; 6A9881232A09766A CRC64;

Query Match 66.9%; Score 716; DB 16; Length 213;
Best Local Similarity 64.8%; Pred. No. 3.9e-50;
Matches 138; Conservative 27; Mismatches 48; Indels 0; Gaps 0;

QY 1 MITIVLLIAYLLGSIPLSGIWLQGVFFQINLREHSGNVTGTTNTRILGKAGMATFVID 60
Db 1 MKLLFITAYLLGSIPTGLWIGQYFHNLREHSGNVTGTTNTRILGVKAGTATLAI 60

QY 61 PFKGLTATLPIIFHLQGVSPILFGLLAVIGHTFIFAGFKGKAVATSAGVIFGAPIF 120
Db 61 MFKGLTALLPIIFGNTSISSAIGFAVLGHTFIFANFKGKAVATSAGVLLGFAPLY 120

QY 121 CLYLAIIFGALYGLSISLSSVTASTAAVIGVLLPFLFGFILSNYSDFIAIILALAS 180
Db 121 LFFLASIFVLVLSMISLSSVASTAAVIGVLSVLTFFPAIHLLPNYDFLTPIVILLAFI 180

QY 181 IIRHKDNIARIKNNKTNLVPWGLNLTHODPKK 213
Db 181 IIRHKDNIARIKNNKTNLVPWGLNLTHODPKK 213

RESULT 3
Q9X972 PRELIMINARY: PRT: 168 AA.
ID Q9X972 AC
AC Q9X972:
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE Hypothetical 17.9 kDa protein (Fragment).
OS Streptococcus gordonii.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
```

```
OX NCBI_TaxID=1302;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CHI;
RA Vriesema A.J., Dankert J., Zaat S.A.;
RT "Isolation and characterization of promoter regions from Streptococcus gordonii CHI.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ236899; CAB40549.1; -;
DR InterPro; IPR003811; DUF205.
DR Pfam; PF02660; DUF205; 1.
KW Hypothetical protein.
FT NON_TER 168 168
SQ SEQUENCE 168 AA; 17921 MW; A07262BD799A478A CRC64;

Query Match 58.0%; Score 621; DB 2; Length 168;
Best Local Similarity 72.1%; Pred. No. 1.3e-42;
Matches 119; Conservative 20; Mismatches 26; Indels 0; Gaps 0;

QY 3 TIVLLIAYLLGSIPLSGIWLQGVFFQINLREHSGNVTGTTNTRILGKAGMATFVIDFF 62
Db 4 TILGLIAYLLGSIPTGLWIGQYFHNLREHSGNVTGTTNTRILGKAGTATLAI 63

QY 63 KGTATLPIIFHLQGVSPILFGLLAVIGHTFIFAGFKGKAVATSAGVIFGAPIFCL 122
Db 64 KGTATLPIIFHLQGVSPILFGLLAVIGHTFIFAGFKGKAVATSAGVIFGAPIFCL 123

QY 123 YLAIIFFGALYGLSISLSSVTASTAAVIGVLLPFLFGFILSNYD 167
Db 124 YLIIFITVLYGLSISLSSVAVAGFAIISVLIPLGILGPSYD 168

RESULT 4
Q8Y7J3 PRELIMINARY: PRT: 198 AA.
ID Q8Y7J3 AC
AC Q8Y7J3:
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein Imol284.
GN Imol284.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Listeriaceae; Listeria.
OX NCBI_TaxID=1639;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-EGD-E / SEROVAR 1/2A;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetoui F., Couve E., de Daruvar A., Dehoux P.,
RA Donann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kretz J., Kuhn M., Kunst F., Kurapat G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordsiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL; AL591978; CAC99362.1; -;
DR ListList; LMO01284; -;
DR InterPro; IPR003811; DUF205.
DR Pfam; PF02660; DUF205; 1.
DR TIGRFAMs; TIGR00023; DUF205; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 198 AA; 21599 MW; 7807B5406DF05CD1 CRC64;

Query Match 43.3%; Score 463.5; DB 16; Length 198;
Best Local Similarity 47.8%; Pred. No. 7e-30;
Matches 96; Conservative 38; Mismatches 56; Indels 11; Gaps 4;
```

```

QY 4 IVLLILAYLGSIPSLGWGVFOINLREHSGNGTGTNTFRILGKAGMATFVIDFFK 63
DB 6 ILLSLAYVIGSIPSLGWIKRIFKKDIRFGSGNLGATNSFRVLGRAGSIVTVMDILK 65

QY 64 GTLTLPLPIIFHLOGVSP--LIFGLLAVIGHTFPIFAGFKGKAVATSAGVIFGPAIFC 121
DB 66 GTVATLLPFFOLNDDHFWLLTGAFALIGHSPFLFAGFRGKAVATSNAYILAYAPLLF 125

QY 122 LYLAIIFGALYGLSMISLSSVTASIAVGVLLPFLGFLNSYDSLFIAIILALASLI 181
DB 126 VAALVFLVTLKSLYKSVLSMIGALALI-ISLF--NG-----DWILVLVACIALFV 176

QY 182 IIRHKDNTARIKNKTENLVLPW 202
DB 177 IWRHRANITRIRNGEPEKIKW 197

RESULT 5
Q92C68 PRELIMINARY: PRT: 198 AA.
AC Q92C68;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical protein lin1323.
GN Lin1323.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Listeriaceae; Listeria.
OX NCBI_TaxID=1642;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CLIP 11262 / SEROVAR 6A;
RX PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoub P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Krest J., Kuhn M., Kunst F., Kutapkat G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schlueter T., Smoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehlend J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL: AL596168; CAC96554.1; -
DR ListList; LIN01323; -
DR InterPro; IPR003811; DUF205.
DR Pfam; PF02660; DUF205; 1.
DR TIGRFAMs; TIGR00023; DUF205; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 198 AA; 21632 MW; B161D1055B203406 CRC64;

Query Match 42.9%; Score 459.5; DB 16; Length 198;
Best Local Similarity 46.0%; Pred. NO. 1.5e-29;
Matches 93; Conservative 41; Mismatches 55; Indels 13; Gaps 4;

QY 4 IVLLILAYLGSIPSLGWGVFOINLREHSGNGTGTNTFRILGKAGMATFVIDFFK 63
DB 6 ILLSLAYVIGSIPSLGWIKRIFKKDIRFGSGNLGATNSFRVLGRAGSIVTVMDILK 65

QY 64 GTLTLPLPIIFHLOGVSP--LIFGLLAVIGHTFPIFAGFKGKAVATSAGVIFGPAIF 120
DB 66 GTVATLLPFFOL-VNHHFWLLTGAFALIGHSPFLFAGFRGKAVATSNAYILAYAPLL 124

QY 121 CLYLAIIFGALYGLSMISLSSVTASIAVGVLLPFLGFLNSYDSLFIAIILALASL 180
DB 125 FVAALVFLTLTKISKYVLSMIGALALI-----ISFPMG--DWILVLVACIALF 175

QY 181 IIRHKDNTARIKNKTENLVLPW 202

```

```

DB 176 IWRHRANITRIRNGEPEKIKW 197

RESULT 6
Q99UC5 PRELIMINARY: PRT: 202 AA.
AC Q99UC5;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical protein SAV1353.
GN SAV1353 OR Sal187.
OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and
OS Staphylococcus aureus (strain N315).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Staphylococcus.
OX NCBI_TaxID=158878; 158879;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.aureus (strain Mu50), and S.aureus (strain N315);
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Ouchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani U. Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus.";
RL Lancet 357:1225-1240(2001).
DR EMBL: AP003362; BAB57515.1; -
DR EMBL: AP003133; BAB42445.1; -
DR InterPro; IPR003811; DUF205.
DR Pfam; PF02660; DUF205; 1.
DR TIGRFAMs; TIGR00023; DUF205; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 202 AA; 22232 MW; A9DA126B5731749C CRC64;

Query Match 41.2%; Score 441.5; DB 16; Length 202;
Best Local Similarity 42.9%; Pred. No. 4.2e-28;
Matches 90; Conservative 43; Mismatches 60; Indels 17; Gaps 3;

QY 1 MITVLLILAYLGSIPSLGWGVFOINLREHSGNGTGTNTFRILGKAGMATFVID 60
DB 1 MMIIIVMLLSYLIGAPPSGVIGKLFKKDIRFGSGNGTATNSFRVLGRPAGFLVTFD 60

QY 61 FFKGTATLLPIIFHLOGVSP-----LIFGLLAVIGHTFPIFAGFKGKAVATSAGV 112
DB 61 IFKGEITVFFFLWLPVHADGPISFTFTNGLLVGLFAILGHVYPVLYLKFGGKAVATSAGV 120

QY 113 IFGPAIFCLYLAIFFGALYGLSMISLSSVTASIAVGVLLPFLGFLNSYDSLFIA 172
DB 121 VLVGNPILLILAIFFIVLAIKFKYVLSASIVRAICCVIGSL-----IIDYILLVVS 173

QY 173 IILADALSIIRHKDNTARIKNKTENLVLPW 202
DB 174 FLVSI--ILIIHRHSNIARIERGEPEKIKW 201

RESULT 7
Q8RFY9 PRELIMINARY: PRT: 194 AA.
AC Q8RFY9;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical membrane-spanning protein FN0537.
GN FN0537.
OS Fusobacterium nucleatum (subsp. nucleatum).
OC Bacteria; Fusobacteria; Fusobacterium.
OX NCBI_TaxID=76856;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 25586;
RX MEDLINE-21886394; PubMed-11889109;
RA Kapral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
RA Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
RA Vasileva O., Chu L., Kogan Y., Chaga O., Goltsman E., Bernal A.,
RA Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,
RA Fonstein M., Kyrplides N., Overbeek R.;
RT "Genome sequence and analysis of the oral bacterium Fusobacterium
RT nucleatum strain ATCC 25586."
RL J. Bacteriol. 184:2005-2018(2002).
DR EMBL: AE010565; AAL94733.1; ...
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 194 AA; 21343 MW; DE327E0AE835915 CRC64;

Query Match 35.9%; Score 385; DB 16; Length 194;
Best Local Similarity 39.4%; Pred. No. 1.4e-23;
Matches 78; Conservative 40; Mismatches 72; Indels 8; Gaps 2;

QY 1 MITIVILLAYLGLSGIPSGLWIGVFFOINLREHSGNGTGTNTFRILGKKGAMATFVID 60
DB 1 MAFFECFIVLTFTYGAIPSGWIGKAFKGVDRDYGKNSGATNSYRVLGAKLGVAVLIND 60
QY 61 FFKGTATLPIIFHLQGVSPILFGLLAVIGHFTPIFAGFKGKAVATSAGVIFGPAIF 120
DB 61 VLKGFIPLYIASKFNVLVNDVLGLVAILAHTFSCFISFKGKGAVATSLGVFLFLPIVI 120
QY 121 CLVLAIFFGALYLGSMISLSSVTASIAAVIGVLLPFLPGFGLSNVDSLFIAIILASL 180
DB 121 TLILLAIFILVATFYKVSLSATIAA-----FLUPIFTF-FTHKDSYLFALSIIAVF 172
QY 181 IIRHKDNIARIKNKTEN 198
DB 173 VIYRHTNLSRLSGTEN 190

RESULT 8
Q8R9J2 PRELIMINARY; PRT; 198 AA.
AC Q8R9J2;
DT 01-JUN-2002 (TRENBLrel. 21, Created)
DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
DE Hypothetical protein TTE1618.
GN TTE1618.
OS Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Thermoanaerobacteriales; Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=119072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MB4T / JCM11007;
RX MEDLINE-21992816; PubMed-11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of T. tengcongensis genome."
RL Genome Res. 12:689-700(2002).
DR EMBL: AE013117; AM24820.1; ...
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 198 AA; 21016 MW; 252C3FF7D512BF02 CRC64;

Query Match 29.1%; Score 311.5; DB 16; Length 198;
Best Local Similarity 37.1%; Pred. No. 1.1e-17;
Matches 73; Conservative 35; Mismatches 80; Indels 9; Gaps 3;

QY 4 IVLLIAYLGLSGIPSGLWIGVFFOINLREHSGNGTGTNTFRILGKKGAMATFVIDFFK 63
DB 4 VLVAIYLLIIGCINNAYIFTKYTRNIDIRNYGSGNAGATNVLRLVLPKAAAPVFLDLVK 63
QY 64 GTLATLPIIFHLQGVSPILFGLLAVIGHFTPIFAGFKGKAVATSAGVIFGPAIFCLY 123
DB 64 GTLATLPIIFHLQGVSPILFGLLAVIGHFTPIFAGFKGKAVATSAGVIFGPAIFCLY 123

Db 64 GVVAVLLGKYP-IGMPGALIAAGIAVVGHNWPIFLKRGKGAVATSVGVVMTINPLGLI 122
QY 124 LAIIFFGALYLGSMISLSSVTASIAAVIGVLLPFLPGFGLSNVDSLFIAIILASLI 183
DB 123 ALAIGVAVIATRYVYSLGSMGTGATFALLNIFFP-----NSVQVLTFAIVLAL--LVIF 174
QY 184 RHKDNIARIKNKTENLV 200
DB 175 QHRSNIKRLINGTESKI 191

RESULT 9
Q9KCD3 PRELIMINARY; PRT; 206 AA.
AC Q9KCD3;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Hypothetical protein BHI639.
GN BHI639.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-C-125 / JCM 9153;
RX MEDLINE-20512582; PubMed-11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis."
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL: AF001512; BAB05358.1; ...
DR InterPro: IPR003811; DUF205.
DR Pfam: PF02660; DUF205; 1.
DR TIGRFAMs: TIGR00023; DUF205; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 206 AA; 21890 MW; 6C7CA6474E4013FA CRC64;

Query Match 27.8%; Score 298; DB 16; Length 206;
Best Local Similarity 33.0%; Pred. No. 1.4e-16;
Matches 65; Conservative 49; Mismatches 75; Indels 8; Gaps 4;

QY 5 VLLIAYLGLSGIPSGLWIGVFFOINLREHSGNGTGTNTFRILGKKGAMATFVIDFFK 64
DB 6 LVVIGSYLLGSVSFYITAKKIKKVDIQRHSGNAGATNLRVLGVGPAVTLLDILK 65
QY 65 TLATLPIIFHLQGVSPILFGLLAVIGHFTPIFAGFKGKAVATSAGVIFGPAIFCL 122
DB 66 VIAVVTVQLTPDGDGWFMAAAGIAAIGHNWPYIYFGRGKGVATTIGVLASLYPLAAV 125
QY 123 YLAITFFGALYLGSMISLSSVT-ASIAAVIGVLLPFLPGFGLSNVDSLFIAIILASLI 181
DB 126 LAGVIAIGSVITRYVYSLGSLFLVTLTALLVAVLQWFGYPVA---YIVLTIVAILSN- 181
QY 182 IIRHKDNIARIKNKTEN 198
DB 182 -WRHRSNIORLLSGTEN 197

RESULT 10
Q9RSV1 PRELIMINARY; PRT; 198 AA.
AC Q9RSV1;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Hypothetical protein DR2021.
GN DR2021.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;
```

```

OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R1;
RA MEDLINE=20036896; PubMed=10567266;
RX White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vanathevan J.J., Lam P., McDonald L., Utterback T., Fleischmann R.D.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1.";
RL Science 286:1571-1577(1999).
DR EMBL; AE002039; AAF11571.1;
DR TIGR; DR2021;
DR InterPro; IPR003811; DUF205.
DR Pfam; PF02660; DUF205; 1.
DR TIGRFAMs; TIGR00023; DUF205; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 198 AA; 20713 MW; 114362BD2E712BC1 CRC64;

Query Match 27.2%; Score 291.5; DB 16; Length 198;
Best Local Similarity 34.5%; Pred. No. 4.6e-16;
Matches 68; Conservative 39; Mismatches 81; Indels 9; Gaps 3;

OY 1 MITIVLLIAYLLGSLIPSGMTQVFFQINLREHSGNNTGTTNFRILGKKGAGMATFVID 60
Db 3 LTALLALLSYLIGAIIPAAMALARA-RGVDIRKVGSGNGATNVLRLSLGKGPALLVASPD 61
OY 61 FPKGLATLPIIFHLOGVSPILFGLLAVIGHTFTPIFAGFKGKAVATSGVIFGFAPIF 120
Db 62 ILKGLVALVLLARALGSAEWAALCGVLAVIGHNFSPLAFRGKGVATSGVFIATLDPV 121
OY 121 CLVLAIFFGALYSGMISLSVTSIAAIVGVLLFPLFGFILSNVDSLFATILALASL 180
Db 122 GLTTFVLATACHWLRFVSAGSINGAFIAGALVLP-----RPTWDR---NAVLF 173
OY 181 IIRHKDNIARIKNKTE 197
Db 174 LVWQHRENIRKLQAGTE 190

RESULT 11
OY Q98M84 PRELIMINARY; PRT; 195 AA.
AC Q98M84;
DT 01-OCT-2001 (TREMBlrel. 18, Created)
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical protein mlr0688.
GN MLR0688.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpō S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti.";
RL DNA Res. 7:331-338(2000).
DR EMBL; AF002995; BAB48229.1;
DR InterPro; IPR003811; DUF205.
DR Pfam; PF02660; DUF205; 1.

DR TIGRFAMs; TIGR00023; DUF205; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 195 AA; 20237 MW; 06FA01BE9DC7BEB9 CRC64;

Query Match 26.8%; Score 286.5; DB 16; Length 195;
Best Local Similarity 36.9%; Pred. No. 1.1e-15;
Matches 73; Conservative 37; Mismatches 75; Indels 13; Gaps 4;

OY 4 IVLLIAYLLGSLIPSGMTQVFFQINLREHSGNNTGTTNFRILGKKGAGMATFVIDFFK 63
Db 6 ILALVFGVLLGSLIPFGLLLTAAAGLDVVRKIGSGNIGATNVLRTGNGKLAATLLDALK 65
OY 64 GTLATLPIIFHLOGVSPILFGLLAVIGHTFTPIFAGFKGKAVATSGVIFGFAPIFCY 123
Db 66 GTAAVL--IAGHFAPETAVWAGLGLHLPVWLGKGGKGVATYGLVGLIGLAQVALI 123
OY 124 LAIIFFGALYSGMISLSVTSIAAIVGVLLFPLFGFILSNVDSLFATILALASLI 182
Db 124 FAVIWLAMAFLEFRYSLLAALTA--VVPVIALYFLS---APOIATVLFVVMSTIIV 173
OY 183 IIRHKDNIARIKNKTENLV 200
Db 174 IHRANISRLLAGTEGKI 191

RESULT 12
OY Q8YC64 PRELIMINARY; PRT; 201 AA.
AC Q8YC64;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Putative integral membrane protein.
GN BMEI0668.
OS Brucella melitensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29459;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=16M / ATCC 23456 / BIOTYPE 1;
RX MEDLINE=20020109; PubMed=11756688;
RA DelVecchio V.G., Kapatral V., Redkar R.J., Patra G., Mujar C., Los T.,
RA Ivanova N., Anderson I., Bhattacharyya A., Lykidis A., Reznik G.,
RA Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goltzman E.,
RA Seikov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
RA Haselkorn R., Kyrpides N., Overbeek R.;
RT "The genome sequence of the facultative intracellular pathogen
RT Brucella melitensis.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
DR EMBL; AE009702; AAL53910.1;
DR InterPro; IPR003811; DUF205.
DR Pfam; PF02660; DUF205; 1.
DR TIGRFAMs; TIGR00023; DUF205; 1.
KW Complete proteome.
SQ SEQUENCE 201 AA; 20507 MW; 105CA44587BB4CA1 CRC64;

Query Match 26.4%; Score 282.5; DB 16; Length 201;
Best Local Similarity 36.6%; Pred. No. 2.5e-15;
Matches 71; Conservative 31; Mismatches 81; Indels 11; Gaps 3;

OY 7 LILAYLLGSLIPSGMTQVFFQINLREHSGNNTGTTNFRILGKKGAGMATFVIDFFK 66
Db 15 LIFGVLGSLIPGLILTRLAGLDVRAIGSGNIGATNVLRTGNGKLAATLLDALKGTA 74
OY 67 ATLPLPIIFHLOGVSPILFGLLAVIGHTFTPIFAGFKGKAVATSGVIFGFAPIFCYLAI 126
Db 75 AAL--IAAHFGQNAATAAGFGAFIGHLPVWLGKGGKGVATYGLVGLIGLAQVALI 132
OY 127 IFFGALYSGMISLSVTSIAAIVGVLLFPLFGFILSNVDSLFATILALASLIIRHK 186
Db 133 AMIVTALLARYSSLSALVASL--VVPVIALY-----SRGNOALALFAIMTVIIFIKR 183

```

```

QY 187 DNIRIKKNTENLV 200
Db 184 ANISRLNGTESKI 197

RESULT 13
Q8UFUL
ID Q8UFUL PRELIMINARY; PRT; 205 AA.
AC Q8UFUL
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical protein Atul306.
GN ATU1306 OR AGF_C_2402.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A. PubMed=11743193;
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Boyce D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutyavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58."
RL Science 294:2317-2323(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houmlel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA Wollam K., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gursion J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58."
RL Science 294:2323-2328(2001).
DR EMBL; AE009092; AAL42312.1; ALT_INIT.
DR EMBL; AE008058; AAK87097.1;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 205 AA; 21678 MW; 973A5024E0CFDE46 CRC64;

Query Match 26.38; Score 282; DB 16; Length 205;
Best Local Similarity 35.18; Pred. No. 2.8e-15;
Matches 68; Conservative 40; Mismatches 74; Indels 12; Gaps 4;

QY 8 ILAYLLGSLPGWIGVFFQINLRHSGNGTGTNTFRILGKAGMATFVIDFKGTLA 67
Db 19 LIGYLLGSLPGFLITRNAGLGDVRKIGSGNIGATNVLRTGNKXLAATLALDKGTAA 78

QY 68 TLLPIIFHLQGV-SPLIFGLLAVIGHTPPIFAGFGKGVATSVAGVIFGFAPICFLYLAI 126
Db 79 VL--VANALWGEASLVAGFFAFGLHPVWLFGFGKGVAVYIGVLGAAPLMLAFAL 136

QY 127 IFFGALYLGSMISLVSSTASIAVIGVLLFPLFGFILSNYDSLFIATILASLIIRHK 186
Db 137 IWLATAFITRYSSLS-----ALLAMLIIPVALWLGPEKTAMLVLLSVISN--WKHR 187

QY 187 DNIRIKKNTENLV 200
Db 188 ENIRLMAGTESRI 201

RESULT 14

```

```

Q97G69
ID Q97G69 PRELIMINARY; PRT; 198 AA.
AC Q97G69;
DT 01-OCT-2001 (TREMBlrel. 18, Created)
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Uncharacterized conserved membrane protein, YgiH/Upp0078
DE family.
GN CAC2500.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Clostridiales; Clostridiaceae; Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Oiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum."
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL; AE007749; AAK80454.1;
DR InterPro; IPR003811; DUF205.
DR Pfam; PF02660; DUF205; 1.
KW Complete proteome.
SQ SEQUENCE 198 AA; 21099 MW; E50FEBDAFF82BBBA CRC64;

Query Match 25.78; Score 275.5; DB 16; Length 198;
Best Local Similarity 32.08; Pred. No. 8.9e-15;
Matches 65; Conservative 39; Mismatches 82; Indels 19; Gaps 4;

QY 1 MITVLILAYLLGSLPGWIGVFFQINLRHSGNGTGTNTFRILGKAGMATFVID 60
Db 2 ILVIITVWVSFLCGSIPTGYLITRKLSGIDVTRKSGNIGTGNVTRVAGTKISMITQMD 61

QY 61 FFKQTALTLPIIFHLQGVSP-----IFGLLAVIGHTPPIFAGFGKGVATSVAGVI 113
Db 62 ILKGIIPVLLCMLIASIKLPISTSMVLSIIIVIAVLGHDTYTPFGNGGKGVNTVGAF 121

QY 114 FGPAIPICFLYLAIFFGALYLGSMISLVSSTASIAVIG--VLLFPLFGFILSNYDSLFI 171
Db 122 FLAPAAVLAVAGAVYFVLRFTKIVSKISAVGITMPTACIALRLPI-----ETV 172

QY 172 AILALASLIIRHKDNIRIKKNTKE 197
Db 173 CALIA-CGLLIRHKDNIRLRLVNEE 197

RESULT 15
Q9JUL4
ID Q9JUL4 PRELIMINARY; PRT; 200 AA.
AC Q9JUL4;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Putative integral membrane protein.
GN NMA1261.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=22491 / SEROGROUP A / SEROTYPE 4A;
RX MEDLINE=20222536; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies S.R., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrell B.G.;

```

RT *Complete DNA sequence of a serogroup A strain of Neisseria

meningitidis 22491.:

RL Nature 404:502-506(2000).

DR EMBL; AL162755; CAB84516.1; -.

DR InterPro; IPR003811; DUF205.

DR Pfam; PF02660; DUF205; 1.

DR TIGRFAMs; TIGR00023; DUF205; 1.

KW Complete proteome.

SO SEQUENCE 200 AA; 20796 MW; DF1BEE95E87C43FF CRC64;

Query Match 25.7%; Score 275.5; DB 16; Length 200;

Best Local Similarity 34.1%; Pred. No. 9e-15;

Matches 70; Conservative 39; Mismatches 82; Indels 15; Gaps 4;

Oy 1 MITIVLLILAYLIGSIFSGLWIGVFFQINLRHGSGNTGTTNTRILGKKAGMATFVID 60

Db 1 MFNIPAVAVSYLIGSLSFVIVSKYYGMDDPRTYSGNPGATNVLRSCKKKAALTLLGD 60

Oy 61 FFKGTATLPLPIFH---LQGVSPILIFGLLAVIGHTPIFAGFKGKAVATSAGVIFGF 116

Db 61 AAKGLVALLARVLOEPLGLSDSAIAAVALAALVGHMWPVFFGFKGKGATALGVLLAL 120

Oy 117 APIFCLYLAIFFGALYLGSMISLSSVTASTAAVIGVLLF-PLFGFILSNYDSLFIAIL 175

Db 121 SPTTALVCALIMLVMAFGFKVSSLAALTATIAAPLAALFFMPHTSWIFAT-----L 171

Oy 176 ALASLIIRHKDNIAR-IKKNTENL 199

Db 172 AIAILVLLRHKSNIILNLIKESKI 196

Search completed: December 26, 2002, 01:16:19

Job time : 689.791 secs

QY 1 MITIVLLILAYLLGSPSLWIGOVFOINLREH-----GSGNTGTTNTFRILGKKGAMAT 56
Db 118 IVCIPAFIQOFGINAVIFSSIFAKAGLGEASILGSGVIGITINVLVTI-----VAI 172
QY 57 FVIDPFK-----GTLATLLPIIFHLOGVSLIFGLLAVIGHTPFIAGFKGKAV 106
Db 173 FVVDKIDRKKLLVGGNIGMIASLL-----IMAILWTI-----GI 207
QY 107 ATSAGVIFGAPICFLYLAIIFFGALY-----LGSMSLSSTASIAAV 150
Db 208 ASSAWII-----IVCLSLFIVFGISGVPVLMVLPFPMRARGAATGISALVNLIGPL 262
QY 151 IGVLLPFLPGFILS-NYDSLFIATILALASLIIR 184
Db 263 IVSLFPILSDALSTEWFLIFAFGLVAMIFVIK 297

RESULT 6

US-09-134-001C-4999
; Sequence 4999, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4999
; LENGTH: 405
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4999

Query Match 8.6%; Score 92.5; DB 4; Length 405;

Best Local Similarity 23.3%; Pred. No. 0.034;
Matches 50; Conservative 31; Mismatches 69; Indels 65; Gaps 12;

QY 42 TNERILGKKGAMATFVIDPFKGLATLPII-----FHL-OGVSLPIFGLLAVIGH 92
Db 21 TKSFNJ-----NFAINFFVLCWYLLVIVIASYKSEYHSDSVAGLVVGLF-IVGS 71
QY 93 TPEIFAGFKGKAVATSAGVIFGAPICFLYLA-IFF--GALYLGSMISL----- 140
Db 72 LIGRFVTGKYVNRFGPKLIFG---LICLVVTQLLYFIPGVSWMVLMVRLNGLATAVA 128
QY 141 SSVTASIAA-----VIGVLLPFLPGFILSNYDSL-----FIAI 173
Db 129 TTATGTIAAYITPPTKSEGISLSLSVLGTAGIPFGFMGLMMSFSINILFTICVILGV 188
QY 174 ILALASLIIRKNDIARIKNTENLVPNGLNTH 208
Db 189 ISGLSLLI---KINFTTVKENT---ITHKRFNLAH 218

RESULT 7

US-09-134-001C-5055
; Sequence 5055, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08

; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5055
; LENGTH: 287
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5055

Query Match 8.5%; Score 91.5; DB 4; Length 287;

Best Local Similarity 23.9%; Pred. No. 0.027;
Matches 60; Conservative 37; Mismatches 87; Indels 67; Gaps 14;

QY 8 ILAYLLGSPSLWIGOVFOINLREHSGN-----TGTTNTFRILGKKGAMATFVIDPFK 53
Db 3 IVDFLIALLPALFWGSVIINVFV---GGGPNQIRGTT-----IGTLFIGF-- 46
QY 64 GTLAT-----LLPIIFHLOGVSPILFGL-----LAVIGHTFPIFAGFKGKAVAT 108
Db 47 SLLATGHAADFNDLTVIIVGLVSGALWAFGQGNQLKSVHLIGVSKTPISTGHLVGTTLF 106
QY 109 SAGVIFGAPICFLYLAIFFGALYLGSMISLSVTA-----SIAAVIGVLLPLF 159
Db 107 SAIFLGENSTIVQVVMGLIAMILLVVG--ISLTSKAKSEKSDNPEFKKANGILLSTI 164
QY 160 GFI-----LSNYDSLFI-AIILALASLIIRKNDIARIKNTENLVP---WGL 204
Db 165 GYGVVVLGDIFGVSQTDALFFQSIGMAIGGLISLNMHNHTS---IKSTALNLPVINGI 221
QY 205 N---LTHODPK 212
Db 222 GNLFMFYSQPK 232

RESULT 8

US-08-677-049-9
; Sequence 9, Application US/08677049
; Patent No. 5858707
; GENERAL INFORMATION:
; APPLICANT: Guimaraes, M. Jorge
; APPLICANT: Bazan, J. Fernando
; APPLICANT: McClanahan, Terrill K.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: PURIFIED MAMMALIAN NUCLEOBASE PERMEASES;
; TITLE OF INVENTION: NUCLEIC ACIDS; ANTIBODIES
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DNAX Research Institute
; STREET: 901 California Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/677,049
; FILING DATE: 03-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/000,788
; FILING DATE: 03-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Chiqo, Edwin P.
; REGISTRATION NUMBER: 34,090
; REFERENCE/DOCKET NUMBER: DX0511
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-852-9196
; TELEFAX: 415-496-1200
; INFORMATION FOR SEQ ID NO: 9;

SEQUENCE CHARACTERISTICS:

LENGTH: 438 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:

NAME/KEY: Region
LOCATION: 115..144
OTHER INFORMATION: /note= "Encompasses TM 4 of Figure"
OTHER INFORMATION: 4

FEATURE:
NAME/KEY: Region
LOCATION: 323..357
OTHER INFORMATION: /note= "Encompasses TM 9 of Figure"
OTHER INFORMATION: 4

FEATURE:
NAME/KEY: Region
LOCATION: 359..386
OTHER INFORMATION: /note= "Encompasses TM 10 of Figure"
OTHER INFORMATION: 4

US-08-677-049-9

Query Match 8.5%; Score 91.5; DB 2; Length 438;

Best Local Similarity 24.9%; Pred. No. 0.048;

Matches 44; Conservative 31; Mismatches 51; Indels 51; Gaps 11;

QY 47 ILGKAGMA---TFV--IDFFKGTATLLPI---IFHLOGVSLIFGLLAVIGHTF-- 94
DB 29 IVKRAMGLTEQLTVLSIDIFMCGVATLLQWNSRFFGI-----GLPVLGCTFTA 80
QY 95 --PIFAGKGGKAVATSGVIFGAPIFCLYLAIIFFGALYLGSMISL--SSVTASIAAV 150
DB 81 VSPMI-----AIGSEGVSTVYGSIIASGILVILI-SFFPGKLVSPFPVVTGVSVTI 132
QY 151 IGVLLPPL-----RGFILSNVDSLFIATLILALASLIIRHKDNRIARKN 194
DB 133 IGITLMPVAMNMGEGSADFG-DLSN-----LALAFVLSIIVLLYRFTKGFIS 183

RESULT 9

US-09-134-001C-4744

; Sequence 4744, Application US/09134001C

; Patent No. 6380370

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS

; FILE REFERENCE: GTC-007

; CURRENT APPLICATION NUMBER: US/09/134,001C

; CURRENT FILING DATE: 1998-08-13

; PRIOR APPLICATION NUMBER: US 60/064,964

; PRIOR FILING DATE: 1997-11-08

; PRIOR APPLICATION NUMBER: US 60/055,779

; PRIOR FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 5674

; SEQ ID NO 4744

; LENGTH: 518

; TYPE: PRT

; ORGANISM: Staphylococcus epidermidis

US-09-134-001C-4744

Query Match 8.5%; Score 91.5; DB 4; Length 518;

Best Local Similarity 23.9%; Pred. No. 0.06;

Matches 39; Conservative 33; Mismatches 72; Indels 19; Gaps 7;

QY 46 RIUGKAGMATFVIDFFKGTATLLPIIFHLOGVSLIFGLLAVIGHTFPFAGF-KGOK 104
DB 117 RVYTEAGDAITLPDFKRLDDKNI-----KIISCLIIIVVFTLYTHSGFVSGK 169
QY 105 AVATSGVIF--GFAPIFCLYLAIIFFGALYLGSMISLSSVTASIAAVIGVLLPFLGFI 162
DB 170 LFESAGLNHAGLLIVAIIVIFTFEGG-YLA--VSITDFQGVIMLIAMVWVPIVALL 226

QY 163 -LSNYSLSFIAIILALASLIIRHKDNRIARKNKTENLVPWGL 204
DB 227 KLINGWDTEHDIAQMKPTNLDLFRGTTVIGIV-----SLFSNGL 264

RESULT 10

US-09-134-001C-2912

; Sequence 2912, Application US/09134001C

; Patent No. 6380370

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCO

; FILE REFERENCE: GTC-007

; CURRENT APPLICATION NUMBER: US/09/134,001C

; CURRENT FILING DATE: 1998-08-13

; PRIOR APPLICATION NUMBER: US 60/064,964

; PRIOR FILING DATE: 1997-11-08

; PRIOR APPLICATION NUMBER: US 60/055,779

; PRIOR FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 5674

; SEQ ID NO 2912

; LENGTH: 400

; TYPE: PRT

; ORGANISM: Staphylococcus epidermidis

US-09-134-001C-2912

Query Match 8.4%; Score 89.5; DB 4; Length 400;

Best Local Similarity 21.8%; Pred. No. 0.071;

Matches 57; Conservative 35; Mismatches 60; Indels 109; Gaps 12;

QY 1 MITIVLLILAYLVGIPSGWIGQVFFOINLREHSGMGTGTTNTPR-----ILGKKA 52
DB 166 MWVFLGIVSLVGLI-----VLVEKKE-----NTNTINKEEKMKNNLNKKF 211
QY 53 GNATFVIDFFKG-----TLATLLPIIFHLOGV-----SPLIFGLLA-- 88
DB 212 VIFFSIATVFFEGAGYIITGTFELVAIVKSIPIVAEYATLSWMFVGLGAIPSTLIWSLIAEK 271
QY 89 -----VIGHTPEIFAGFKGKAVATSGVIFGAPIFCLYLAIIFFGALYL 134
DB 272 ISYKKAIFYAFILQIISVCLPVT-----HEIFSLVISSVLFEGTFL 313
QY 135 G-----SMISLSSVTASIAAVIGVLLPFLPFGFIFL-----SNYD-SL 169
DB 314 GLTTLFISKQSLMYKTDDBPLNVLSTLTVYS-----LGQMLAPMFAGILIGKSNMYIAL 369
QY 170 FTATILALASLI--IIRKDN 188
DB 370 IFATVLLILGLISTIFS YKQN 390

RESULT 11

US-09-134-001C-3501

; Sequence 3501, Application US/09134001C

; Patent No. 6380370

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCO

; FILE REFERENCE: GTC-007

; CURRENT APPLICATION NUMBER: US/09/134,001C

; CURRENT FILING DATE: 1998-08-13

; PRIOR APPLICATION NUMBER: US 60/064,964

; PRIOR FILING DATE: 1997-11-08

; PRIOR APPLICATION NUMBER: US 60/055,779

; PRIOR FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 5674

; SEQ ID NO 3501

; LENGTH: 434

; TYPE: PRT

; ORGANISM: Staphylococcus epidermidis

US-09-134-001C-3501

Query Match 8.2%; Score 88; DB 4; Length 454;
Best Local Similarity 23.3%; Pred. No. 0.12;
Matches 60; Conservative 33; Mismatches 81; Indels 84; Gaps 13;

QY 13 LGSIPSLWTI--GOVFOINLRHSGNGTNTFRILGKKGAMATFVIDF-----FKGTL 66
DB 15 IGSMLGFLPGAGNLFFPHLGOTAGSNWTAN-----LG-----FLTAICLPLGII 63
QY 67 A---TLPIIFHLOGVSLIFGLLAVIGHTF---PIFA-----GFKGKAVATAGVI 113
DB 64 AIGVSKTNGVFEISSRSKIYGLFTIGLVIGPFPALPRLATTSFEIAFSPFISSGTA 123
QY 114 FGPAFICLYLAIFGALYLGSMISLSSVTASIAAIVGVLLPLF----- 159
DB 124 QALLPIF---SILFFGAVL-----FSRKPISKILDYIGKFLNPVFLILGIVVVLAFIR 174
QY 160 -----GFILSNYDSLFAITAILALASLII-----IRHKDNIAIR 191
DB 175 PMGISHAPVSADYSNVLKGF-DGYNTLDALASLAFGIIIVTTIKKLGITNPNTIAK 233
QY 192 --INKNTENLVPWGLNIT 207
DB 234 ETLKSGTISIAMGVYIT 251

RESULT 12

US-09-165-396-4

; Sequence 4, Application US/09165396

; Patent No. 6441134

; GENERAL INFORMATION:

; APPLICANT: BECKER, JEFFREY W.

; APPLICANT: LUBKOWITZ, MARK A.

; TITLE OF INVENTION: ISOLATED CANDIDA ALBICANS OLIGOPEPTIDE TRANSPORTER GENE

; FILE REFERENCE: 372.6520P

; CURRENT APPLICATION NUMBER: US/09/165,396

; CURRENT FILING DATE: 1998-10-02

; EARLIER APPLICATION NUMBER: PCT/US98/02332

; EARLIER FILING DATE: 1998-02-06

; EARLIER APPLICATION NUMBER: 60/037,859

; EARLIER FILING DATE: 1997-02-07

; NUMBER OF SEQ ID NOS: 15

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 4

; LENGTH: 799

; TYPE: PRT

; ORGANISM: S. cerevisiae

US-09-165-396-4

Query Match 8.2%; Score 87.5; DB 4; Length 799;
Best Local Similarity 21.9%; Pred. No. 0.3;
Matches 46; Conservative 37; Mismatches 82; Indels 45; Gaps 9;

QY 4 IVLLIAYLLGSIPSLGWLIGOVFOINLRHSGNGTNTFRILGKKGAMATFVIDFFK 63
DB 279 LIVLIGSFIWVPGFLTGLSYFVLWLGSKTRHNTANT--IFGTQSLGA----- 329
QY 64 GTLATLLPIIFHLOGVSLIFGLLAVIGHTFPIFAGKGGKAVATAGVIFGFAPIF-CL 122
DB 330 -----LPITFDYTVQVQAMSG--SVFATPVVSAN-----TYASVLIFFVIVLPCL 373
QY 123 YLAIIFFGALYL-----GSMISLSSVTASIAAIVG-----VLLPFLGFIILSN 165
DB 374 YFTNTWY-AKYPVITSGSTYDNTQNKYNVTIKLNEDYSINLEKYKEYSPVFPFVSYLLS- 431
QY 166 YDSLFAITAILALASLIIIRHKDNIAIRKNK 195
DB 432 YALNFAAVIAVFEVHCILTHGKDIVAKFDR 461

RESULT 13

US-09-245-808-1

; Sequence 1, Application US/09245808
; Patent No. 6313277

; GENERAL INFORMATION:

; APPLICANT: Doyle, L. Austin

; APPLICANT: Abruzzo, Lynne V.

; APPLICANT: Ross, Douglas D.

; TITLE OF INVENTION: Breast Cancer Resistance Protein (BCRP) and DNA which

; FILE REFERENCE: encodes it

; CURRENT APPLICATION NUMBER: US/09/245,808

; CURRENT FILING DATE: 1999-02-05

; EARLIER APPLICATION NUMBER: 60/073763

; EARLIER FILING DATE: 1998-02-05

; NUMBER OF SEQ ID NOS: 7

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 1

; LENGTH: 655

; TYPE: PRT

; ORGANISM: Human MCF-7/AdrVp cells

US-09-245-808-1

Query Match 8.0%; Score 86; DB 4; Length 655;
Best Local Similarity 19.2%; Pred. No. 0.34; 75; Indels 80; Gaps 12;
Matches 47; Conservative 43; Mismatches 75; Indels 80; Gaps 12;

QY 9 LAYLLGSIPSLGWLIGOVFOINLRHSGNGTG-----TNTFRILGKKGAMATFVID-- 60
DB 396 IAOIIVTVVLGLVIGLFGKLNKNDSTGIONRAGVLFLLTN--QCFSSVSASVELFVVEKK 453
QY 61 -----FFKGTL-----ATLLP-----IIFHLOGVSP-----LIFGL 86
DB 454 LFIHEITSGYRVSSVFLGKLLSDLLPMTMLPSIIFTCIVYFMLGKPKADAFVMMFTL 513
QY 87 LAVIGHTFPFAGKGGKAVATAGVIFGFAPIFCLYLAIIFFGALYLGSMISLSSVTAS 146
DB 514 MMVAYSASSMALAIAAGSVSVATLLM-----TICFVFMIFSGLL-----VNLTTI-AS 563
QY 147 IAAVIGVLLPFLGFIILSNYDSLFAITAILALASLIIIRHKDNIAIRKNKTNLVPWGLN 206
DB 564 WLSWLOVFSIPRYGF-----TALQHNFEFG-----QNFCEP-CLNA 597
QY 207 THODP 211
DB 598 TGNNP 602

RESULT 14

US-09-134-001C-5389

; Sequence 5389, Application US/09134001C

; Patent No. 6380370

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOX

; FILE REFERENCE: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: GTC-007

; CURRENT APPLICATION NUMBER: US/09/134,001C

; CURRENT FILING DATE: 1998-08-13

; PRIOR APPLICATION NUMBER: US 60/064,964

; PRIOR FILING DATE: 1997-11-08

; PRIOR APPLICATION NUMBER: US 60/055,779

; PRIOR FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 5674

; SEQ ID NO 5389

; LENGTH: 901

; TYPE: PRT

; ORGANISM: Staphylococcus epidermidis

US-09-134-001C-5389

Query Match 8.0%; Score 86; DB 4; Length 901;

Best Local Similarity 24.0%; Pred. No. 0.52; 72; Indels 62; Gaps 10;

Matches 52; Conservative 31; Mismatches 72; Indels 62; Gaps 10;

QY 28 QINLRHSGNGTGTTNTFRILGKKGAMATFVIDFFKGT-ATLLPIIFHLOGVSLIFGL 86

```
Db 169 QIETGGAMNSEPGTSEIVGIVARVILLITF--GSLIAGMPIISAIGLSSV-GI 225
Oy 87 LAVIGHTFPI-----PAGEKGG-----KAVATSAGVIFG 115
Db 226 IALTYIFDIPNFTLTAVMIGLAVGIDYSLFILFRKELKKGVDVTEAIAVGTAGS 285
Oy 116 FAPIFCULAIIFGALYLGSMISLSSVTASIAAVIGVLLPFLPGFILSNYDSLFAIIL 175
Db 286 -----AVIFAGLTVMIAGCLSLVGIDFLAVMG-----FASAISVLFVLAALTL 330
Oy 176 ALASLIIRHKNIARKNKNTLNPWGLNLTHODPK 212
Db 331 -LPALISIPHK--IKIDKPTK-----SKDKP 355
```

RESULT 15

```
US-09-134-001C-5000
; Sequence 5000, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5000
; LENGTH: 190
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5000
```

```
Query Match      8.0%; Score 85.5; DB 4; Length 190;
Best Local Similarity 22.9%; Pred. NO. 0.074;
Matches 47; Conservative 28; Mismatches 69; Indels 61; Gaps 9;

Oy 1 MITIVLLIAYLLGSLGWLIGQVFFQINLREHSGNLTGTTNFRILGKKGAGMATFVID 60
Db 19 LMTAICILG-LVPSVPLPFPVPLVQ-----NIGIFLAGIILGRKLG----- 62
Oy 61 FPKGTATLLPIIFHLQGVSLIFGLLAVIGHTFPFIFAGKGGKAV--ATSAGVIFGPAP 118
Db 63 -----TSVIVFLLVATG--LPVLSGGRGGIGVFPAGPSAGFLFLY-P 101
Oy 119 IFCLYLAIIFGALYLGSMISLSSVTASIAAVIGVLLPFLPGFILSNY-----DSLFI- 171
Db 102 VVAVFIGIIRDAYLHKINFLVFIAF-----LVIGVLGDLILGTLINGFIHPIPSKAEIL 157
Oy 172 -----AIIALASLIIRH 185
Db 158 SFTMPGDIINAIATASLIGAILNH 182
```

Search completed: December 26, 2002, 02:47:12
Job time : 134.781 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: December 24, 2002, 22:13:55 ; Search time 11247.8 Seconds
(without alignments)
1661.118 Million cell updates/sec

Title: US-10-068-080-2
Perfect score: 642
Sequence: 1 atgattacaatagttttatt.....atcaagatcctaaaaataa 642

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rod.*
- 36: em_htg_mam.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	642	100.0	6171	6	BD003717	BD003717 Polynucle
2	642	100.0	10624	1	AE007390	AE007390 Streptococ
3	640.4	99.8	6812	1	SPPARCETP	267739 Streptococ
4	640.4	99.8	10828	1	AE008451	AE008451 Streptococ
5	637.2	99.3	248254	2	SPNEU1903	AL44925 Streptococ
6	388.4	44.9	25021	1	AB028896	AB028896 Streptococ
7	275.6	42.9	11545	1	AE006539	AE006539 Streptococ
8	275.6	42.9	11573	1	AE010024	AE010024 Streptococ
9	274	42.7	53354	1	AE014149	AE014149 Streptococ
10	246.6	38.4	13211	1	AE006332	AE006332 Lactococc
11	244.6	38.1	771	1	SGO236899	AJ236899 Streptococ
12	153.4	23.9	250050	1	AL591978	AL591978 Listeria
13	147	22.9	7563	1	AF084044	AF084044 Listeria
14	124	19.3	195269	6	AX417035	AX417035 Sequence
15	124	19.3	349980	6	AX417044	AX417044 Sequence
16	103.8	16.2	9839	1	AE010565	AE010565 Fusobacte
17	102.4	16.0	3014	1	AF269800	AF269800 Staphyloc
18	102.4	16.0	3014	6	AX145118	AX145118 Sequence
19	102.4	16.0	4105	1	AF269437	AF269437 Staphyloc
20	102.4	16.0	4105	6	AX144757	AX144757 Sequence
21	100.8	15.7	295350	1	AP004826	AP004826 Staphyloc
22	100.8	15.7	303750	1	AP003133	AP003133 Staphyloc
23	100.8	15.7	346900	1	AP003362	AP003362 Staphyloc
24	100	15.6	6595	1	AF024713	AF024713 Bacillus
25	100	15.6	26170	1	BC170DEGR	273234 B.subtilis
26	100	15.6	233780	1	BSUB0010	299113 Bacillus su
27	96	15.0	594	6	AX144133	AX144133 Sequence
28	69.4	10.8	10530	1	AE013956	AE013956 Yersinia
29	68.2	10.6	11198	1	AE004138	AE004138 Vibrio ch
30	67.4	10.5	10177	1	U32713	U32713 Haemophilus
31	65	10.3	7218	6	I66494	I66494 Sequence 14
32	65.2	10.2	300550	1	AP001512	AP001512 Bacillus
33	61	9.5	1033	1	LLRXNRDEF	X92690 L.lactis nr
34	61	9.5	12043	1	AE009702	AE009702 Brucella
35	59	9.2	12422	1	AE004494	AE004494 Pseudomon
36	58.4	9.1	20604	1	AE008847	AE008847 Salmonell
37	57.3	9.0	258050	1	AL627278	AL627278 Salmonell
38	56.5	8.8	346897	1	AP002995	AP002995 Mesorhizo
39	56.4	8.8	14651	1	AE012788	AE012788 Chlorobiu
40	56.2	8.8	8634	1	AF281816	AF281816 Acholepla
41	55.2	8.6	10029	1	AE005535	AE005535 Escherich
42	55.2	8.6	307962	1	AP002564	AP002564 Escherich
43	54.8	8.5	15663	1	AE001796	AE001796 Thermotog
44	53.6	8.3	1641	1	ECOBACTRA	L12966 Escherichia
45	53.6	8.3	11307	1	AE000387	AE000387 Escherich

ALIGNMENTS

RESULT 1
BD003717
LOCUS BD003717 6171 bp DNA linear PAT 31-JAN-2002
DEFINITION Polynucleotide of Streptococcus pneumoniae and sequence.
ACCESSION BD003717
VERSION BD003717.1 GI:18631678
KEYWORDS JP 2001501833-A/37.
SOURCE unidentified.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 6171)
AUTHORS Kunsch,C.A., Choi,G.H., Dillon,P.J., Rosen,C.A., Bara,S.C.,
Fannon,M. and Dougherty,B.A.
TITLE Polynucleotide of Streptococcus pneumoniae and sequence
JOURNAL Patent: JP 2001501833-A 37 13-FEB-2001;

CDS																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																				
-----	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

KPHRVFLAGAAVEQFVGDFFSTVDVNDOTLLVQSNKGTIYESRLQDLNLSERTSNG
SFISDITSDREVFAYLQEVVTEDEK*

BASE COUNT 2106 a 1318 c 1555 g 1833 t

ORIGIN

Query Match 99.8%; Score 640.4; DB 1; Length 6812;
Best Local Similarity 99.8%; Pred. No. 1.7e-150;
Matches 641; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 ATGATTACAATAGTATTTTAAATCTAGCCTATCTGCTGGGTTTCAGTCCATCTCGTCTC 60
DB 1119 ATGATTACAATAGTATTTTAAATCTAGCCTATCTGCTGGGTTTCAGTCCATCTCGTCTC 1060
OY 61 TCGATTGGACAAGTATCTTTCAATCAATCTACGCGACATGTTCTGGTAACACTGGA 120
DB 1059 TCGATTGGACAAGTATCTTTCAATCAATCTACGCGACATGTTCTGGTAACACTGGA 1000
OY 121 AGCACCACACCTTCGCGATTTTAGTAAAGAGCTGGTATGGCAACCTTTCTGATTGAC 180
DB 999 AGCACCACACCTTCGCGATTTTAGTAAAGAGCTGGTATGGCAACCTTTCTGATTGAC 940
OY 181 TTTTCAAGGAACCTTACGACGCTGCTTCCGATTTATTTTCATCTACAAAGCGTTTCT 240
DB 939 TTTTCAAGGAACCTTACGACGCTGCTTCCGATTTATTTTCATCTACAAAGCGTTTCT 880
OY 241 CCTCTCATCTTTGGACTTTTGGCTGTTATCGGCCATACCTCCCTATCTTTTCAGGATT 300
DB 879 CCTCTCATCTTTGGACTTTTGGCTGTTATCGGCCATACCTCCCTATCTTTTCAGGATT 820
OY 301 AAAGTGGTAAAGCTGTCGCAACCACTGCTGGAGTGATTTTCGGATTTTCGCGCTATCTTC 360
DB 819 AAAGTGGTAAAGCTGTCGCAACCACTGCTGGAGTGATTTTCGGATTTTCGCGCTATCTTC 760
OY 361 TGTCTCTACCTTGGCGATTAATCTTTTGGAGCTCTCTATCTTGGCAGTAGATTTTCAC 420
DB 759 TGTCTCTACCTTGGCGATTAATCTTTTGGAGCTCTCTATCTTGGCAGTAGATTTTCAC 700
OY 421 TCTAGTGTACAGCATCGATTCGGCTGTTATCGGGTCTGCTCTTCCACCTTTTGGT 480
DB 699 TCTAGTGTACAGCATCGATTCGGCTGTTATCGGGTCTGCTCTTCCACCTTTTGGT 640
OY 481 TTTATCTCTGAGTAACATGACCTCTCTCTCATCGCTATTTATCTTAGCAGCTTCTGATTTG 540
DB 639 TTTATCTCTGAGTAACATGACCTCTCTCTCATCGCTATTTATCTTAGCAGCTTCTGATTTG 580
OY 541 ATTATCATCTGCTAAGACATATAGCTGATCAAAATAAATAAATAAATTGGTC 600
DB 579 ATTATCATCTGCTAAGACATATAGCTGATCAAAATAAATAAATAAATTGGTC 520
OY 601 CCTTGGGATTTGAACCTAACCTAACCTAACCTAACCTAACCTAACCTAACCTAACCTAAC 642
DB 519 CCTTGGGATTTGAACCTAACCTAACCTAACCTAACCTAACCTAACCTAACCTAACCTAAC 478

RESULT 4

AE008451/c

LOCUS 10828 bp DNA linear BCT 13-SEP-2001
DEFINITION Streptococcus pneumoniae R6 section 67 of 184 of the complete genome.

ACCESSION AE008451 AE007317

VERSION AE008451.1 GI:15458341

KEYWORDS Streptococcus pneumoniae R6.

SOURCE Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

ORGANISM Streptococcus.

REFERENCE 1 (bases 1 to 10828)

AUTHORS

Hoskins, J.A., Alborn, W. Jr., Arnold, J., Blaszcak, L., Burgett, S.,
DeHoff, B.S., Estrem, S., Fritz, L., Fu, D.-J., Fuller, W., Geringer, C.,
Gilmour, R., Glass, J.S., Khoja, H., Kraft, A., Lagace, R.,
LeBlanc, D.J., Lee, L.N., Lefkowitz, E.J., Lu, J., Matsushima, P.,
McAhen, S., McHenry, M., McLeaster, K., Mundy, C., Nicas, T.I.,
Norris, F.H., O'Gara, M., Peery, R., Robertson, G.T., Rockey, P.,

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

Sun, P.-M., Winkler, M.E., Yang, Y., Young-Bellido, M., Zhao, G.,
Zook, C., Baltz, R.H., Jaskunas, S.R., Rosteck, P.R. Jr., Skatrud, P.L.
and Glass, J.I.
Genome of the bacterium Streptococcus pneumoniae strain R6
J. Bacteriol. 183 (19), 5709-5717 (2001)
21429245
11544234
2 (bases 1 to 10828)
Hoskins, J.A., Alborn, W. Jr., Arnold, J., Blaszcak, L., Burgett, S.,
DeHoff, B.S., Estrem, S., Fritz, L., Fu, D.-J., Fuller, W., Geringer, C.,
Gilmour, R., Glass, J.S., Hann, A., Khoja, H., Kraft, A., Lagace, R.,
LeBlanc, D.J., Lee, L.N., Lefkowitz, E.J., Lu, J., Matsushima, P.,
McAhen, S., McHenry, M., McLeaster, K., Mundy, C., Nicas, T.I.,
Norris, F.H., O'Gara, M., Peery, R., Robertson, G.T., Rockey, P.,
Sun, P.-M., Winkler, M.E., Yang, Y., Young-Bellido, M., Zhao, G.,
Zook, C., Baltz, R.H., Jaskunas, S.R., Rosteck, P.R. Jr., Skatrud, P.L.
and Glass, J.I.
Direct Submission
Submitted (27-JUL-2001) Infectious Diseases Research, Eli Lilly and
Company, Lilly Research Labs, Indianapolis, IN 46285-0438, USA
Location/Qualifiers
1. 10828
/organism="Streptococcus pneumoniae R6"
/strain="R6"
/db_xref="taxon:171101"
103..1638
/gene="ABC-NBD"
/note="spr0748"
103..1638
/gene="ABC-NBD"
/note="Putative"
/codon_start=1
/transl_table=11
/label="spr0748"
/product="ABC transporter ATP-binding protein -
ribose/galactose transport"
/protein_id="AAK99552.1"
/db_xref="GI:15458342"
/translation="MAHENVEMRDITKVFQGVANDKINLHKKGEIHALGENGAG
KSLMNLKGLLEPTSGEIVAGVQVNDLSPKASGUGVHGFHMFVAFVVAENI
ITLSELTKGVLDIAGASKKALSERGLAVDPFSAKVADISVGAQVRELTKLYRRG
ADILFDEPTAVLTPSEIDELMAIMKLVKEGKSIILITHRLDEIVRSDVRLTVIRG
KSIETVEIAGTANADLAEMVGRSVFKFQKQKPEVLSIKDLVNVNRRGVPAVK
NLSLDVRAGEIVGIAGIDNGOSELIOATGLRKVESGIELKDSIVGLHPRITEL
SYGVHPEDRHRDGLILEMISENIALOTYYKEPKNGKILNYSITSVAKKLMEEFDV
RAASLVPAAALSGNQKAIAREIDRDPDLIVSQPTRGDVGAIETIHKRLIEER
DNGKAVLVSPFELDEILNVSDRIAVIHDGKIQGVISVPTTNKQGLVLMAGNLGKEK
SDV"
1604..2689
/gene="ABC-MSP"
/note="spr0749"
1604..2689
/gene="ABC-MSP"
/note="Putative"
/codon_start=1
/transl_table=11
/label="spr0749"
/product="ABC transporter membrane-spanning permease -
ribose/galactose transport"
/protein_id="AAK99553.1"
/db_xref="GI:15458343"
/translation="METWERRRVMKSKLQIISVPLISVFLGILLGAIWVWTFGYDAIW
GYELFYTAFLSGLRGIGEIFRANGPLVILGIFAVASRAGFNVGLPQCALACHILSC
WFAUSHPDMPLMILATIVIALAGGIVGAIPGLIRAYLGTSEVIVITMMNITILYV
GNATFHAFPKDFMQSDSTIRVGANATQTPMLTGLTNSRNHIGIFAITAVAVWF
MLKKTTLGFEIRAGNPLHASEYISAKRTIILSMIISGALAGGAGVGLGTQFNY
VVOGSSLAIFNGMAVSLAANSPIGLFAAFLEGLVQVAGPMGMAAQVPSLVSIIVT
ASIIFFVSVHYLLIERFVPRKKQVKGK"
2692..3648
/gene="ABC-MSP"
/note="spr0750"
2692..3648
/gene="ABC-MSP"

FEATURES
source

gene
CDS
gene
CDS
gene
CDS
gene
CDS

/note="Putative"
/codon_start=1
/transl_table=11
/label_spr0750
/product="ABC transporter membrane-spanning permease -
ribose/galactose transport"
/protein_id="AAK99554.1"
/db_xref="GI:15458344"
/translation="MSIITLLPLLVSSMLIYSAPLIPTSGGVFSRGGVVNVVGLGEI
MVMGAFSGVFNLEFAEQFAGAPWMLLVAGLVGSVFSIIHAAATYHFRADHWVSGT
VNLMPALAVFLVKVYNGQTDNLSTQTFGRDFPVLANIPIVIGDIFFKSTLLGYL
AIASFIAWFLTKFTQGLRLRSVGEHPOAADTLGINVYKRWYLGVIISGFLAGIGCA
IYAGISVNSVTVIGPGFIALAAMIFGKNPIGAMLSLFLFGLSGLAVIGSQLPF
LQGPAYVLIQAPIVLTILVLAFFGKAVAPKADGNIIRKSK"
3733..3945
/gene="IS861-truncation"
/note="spr0751"
3733..3945
/gene="IS861-truncation"
/codon_start=1
/transl_table=11
/label_spr0751
/product="Degenerate transposase (orf1)"
/protein_id="AAK99555.1"
/db_xref="GI:15458345"
/translation="MKLSYDVKQIYELRQGSLEKLSNKGFINNFIYMIKLIDR
YGIEFVAKGNRYSPDLQKQEMTHKV"
4081..4248
/gene="IS861-truncation"
/note="spr0752"
4081..4248
/gene="IS861-truncation"
/codon_start=1
/transl_table=11
/label_spr0752
/product="Degenerate transposase (orf1)"
/protein_id="AAK99556.1"
/db_xref="GI:15458346"
/translation="MGRNPKTRPEERTELERLOAENEYLAENAILKKLRELRLKEK
EKERQKLFKN"
4248..4616
/gene="IS861-truncation"
/note="spr0753"
4248..4616
/gene="IS861-truncation"
/codon_start=1
/transl_table=11
/label_spr0753
/product="Degenerate transposase (orf2)"
/protein_id="AAK99557.1"
/db_xref="GI:15458347"
/translation="MTEFSLDLLKAIKLARSTYYYHLKOLDKTDKDOELKAEIQSIF
IEHKNYARRIYLELRNGYLVNHRKRVQHLWKYSIYKLRDRNENILLIKETLARKQ
RISFKNLKALKOWNSATQ"
4592..5080
/gene="IS861-truncation"
/note="spr0754"
4592..5080
/gene="IS861-truncation"
/codon_start=1
/transl_table=11
/label_spr0754
/product="Degenerate transposase (orf2)"
/protein_id="AAK99558.1"
/db_xref="GI:15458348"
/translation="MEOCYDVTVEFAIPVSTOKLYLSPVLDGFENSEIIAYNLSTSPNL
EOVTMLQEAFTKHYENTILHSDQGWQYODSVHRELESKGIQASNRKNSQDNMG
MESFPGILKSEMFVGYEYKFKSLNQLQEAIDYIDYNNRRIKVKLGKLSPVQIRK
FG"
complement(5123..5764)
/gene="spr0755"
complement(5123..5764)
/gene="spr0755"

/note="Putative lipoprotein precursor"
/codon_start=1
/transl_table=11
/label_spr0755
/product="Conserved hypothetical protein"
/protein_id="AAK99559.1"
/db_xref="GI:15458349"
/translation="MITIVLLIILAYLGSIPSGMLIGQVFFQINLREHSGNGTGTNT
FRILGKAGMATEVIDFFKGLTALPLIIFHLQGVSPILFGLLAVIHTFPIFAGFKG
KRAVATSAAGVIFGFAPIFCLYLAIIFFGALYLGSMISLSVSTASIAAVIGVLLPLFG
FILNSYDSFLIAIILASLIIRIHKDNIARIKNKTENLVPMLNLTHODPKK"
5900..7843
/gene="parE"
/note="spr0756"
5900..7843
/gene="parE"
/EC_number="5.99.1.1"
/codon_start=1
/transl_table=11
/label_spr0756
/product="Topoisomerase IV subunit B"
/protein_id="AAK99560.1"
/db_xref="GI:15458350"
/translation="MSKKEININYNDDAIOVLEGLDAVKKRPGMVLGSDGAGLHL
WIEIVNADVEALSGRGDRIDVTINKDGLTVQDHGRGMPGTHMAGIPIVEVITIL
HAGKFTGGQGYRTSGGLHGVSSVWNAISWEIEITRDGAVYKQFENGKRPVTLK
KIGTALKSTGKTVTPMDATIFSTDFKYNITISERLNEFAFLKKNVTLSLTKRDTDE
AIEFHYENGQVDFVSYLNEDKEILTPVLYPEGEDNGFOVEALQYNDGFSNLSVFN
NVRTKDGTHETGLKSAITKVMNDYARKTKLKEKDKNLEGSYDREGLAAVLSILVPE
EHLQFEGOTKDKLGSPLARPVDGIADVLTFFLMENGELASNLIRKAIKARDAREAA
RKARDSENGKKNKDKGLLSGLKLTPAOSKNPAKNELYLVGDSAGSAGKQGRDRFQ
AIPLRGKVINTAKMADILNEEINTMIYITGAGVADFSIEDANYDKIILMTDAD
TDGAHOTLTLTFYRMPRLVEAGHYIALPPLYKMSKKGKKEVAVAMTQGELEE
LRQFGKGATLQRYKGLGEMNADQLWETTPNPETRLIRVTIEDLARAERRVNLMD
KVEPRKVEDNVKFTLEEAIVE"
8254..10734
/gene="parC"
Query Match 99.88; Score 640.4; DB 1; Length 10828;
Best Local Similarity 99.88; Pred. No. 1.7e-150;
Matches 641; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 ATGATTACAATAGTATTTTAAATCCTAGCCTATCTGCTGGTTCGATTCATCTGGTCTC 60
|||||
Db 5764 ATGATTACAATAGTATTTTAAATCCTAGCCTATCTGCTGGTTCGATTCATCTGGTCTC 5705
Qy 61 TGGATTGGACAAGTATTTCTTTCAATCAATCTACGCGAGCATGTTCTGGTAAACACTGGA 120
|||||
Db 5704 TGGATTGGACAAGTATTTCTTTCAATCAATCTACGCGAGCATGTTCTGGTAAACACTGGA 5645
Qy 121 ACGACCAACACCTTCCCGCATTTTAGGTAAGAAAGCTGGTATGGCAACCTTTGTGATTGAC 180
|||||
Db 5644 ACGACCAACACCTTCCCGCATTTTAGGTAAGAAAGCTGGTATGGCAACCTTTGTGATTGAC 5585
Qy 181 TTTTTCAAAGGAACCCCTAGCAACGCTGCTTCCGATTTATTTTCATCTACAGGCGTTTCT 240
|||||
Db 5584 TTTTTCAAAGGAACCCCTAGCAACGCTGCTTCCGATTTATTTTCATCTACAGGCGTTTCT 5525
Qy 241 CCTCTCATCTTTGGACTTTTGGCTGTTTATGGCCATACCTTCCCTATCTTTTGCAGGATTT 300
|||||
Db 5524 CCTCTCATCTTTGGACTTTTGGCTGTTTATGGCCATACCTTCCCTATCTTTTGCAGGATTT 5465
Qy 301 AAAGGTGGTAAAGGCTGCGCAACAGCTGCTGGAGTATTTTCGGATTTGCGCCTATCTTTC 360
|||||
Db 5464 AAAGGTGGTAAAGGCTGCGCAACAGCTGCTGGAGTATTTTCGGATTTGCGCCTATCTTTC 5405
Qy 361 TGTCTCTACCTTGGCATATCTCTTTTGGAGCTCTCTATCTTTGGCAGTATGATTTCCTG 420
|||||
Db 5404 TGTCTCTACCTTGGCATATCTCTTTTGGAGCTCTCTATCTTTGGCAGTATGATTTCCTG 5345
Qy 421 TCTAGTGTTCACAGCATCGATTTCGGCTGTTATCGGGGTCTTATCGGGGTCTTGTCTTTTCCACT 480
|||||
Db 5344 TCTAGTGTTCACAGCATCGATTTCGGCTGTTATCGGGGTCTTATCGGGGTCTTGTCTTTTCCACT 5285


```

/codon_start=1
/transl_table=11
/product="cpsiaa"
/protein_id="BAB62877.1"
/db_xref="GI:15186710"
/translation="DHVDTWKFKELKESHPVLLAFSEHSDSLTSKSLAECLQLTD
LDEYDTNHFYDTNKLKILKILE"
/complement(400..1323)
/gene="cpsiaa"
CDS
  complement(400..1323)
  /gene="cpsiaa"
/function="LysR-type transcriptional regulator"
/codon_start=1
/transl_table=11
/product="cpsiaa"
/protein_id="BAB62878.1"
/db_xref="GI:15186711"
/translation="MRIQLOQYVKKIVETGSMNEAKQLVITOPSLSNVARNLETEMG
IQIRPKGITLKGMFEFLSTARQILEOTALLEERYKGDNTSRELFVSSSHAFV
VNAFVALFNGDVTQYELFRETNRTEIIDVKNRSEIGLVFLNSYNRDLTKLFDD
NSLIATFTTTHPIFVSKSNPLANRKKLSMKDLEDYPYLSYDQGLHNSFYFSEEMMS
QIHPKSVIVSDRATFLNLMIGLDGYVATGILNSKLNKGDEIVAIPLDVDDVDIVVI
RHDKANLSKMGQFIDYLLLEVSFNHKNKOS"
/complement(1447..1452)
-10_signal
  /evidence=experimental
-35_signal
  1453..1457
  /evidence=experimental
-35_signal
  complement(1470..1475)
  /evidence=experimental
-10_signal
  1478..1482
  /evidence=experimental
-10_signal
  complement(1481..1486)
  /evidence=experimental
-35_signal
  complement(1507..1512)
  /evidence=experimental
  1512..1519
  /gene="cpsiaa"
  1512..1519
  /gene="cpsiaa"
/function="regulator"
/codon_start=1
/transl_table=11
/product="cpsiaa"
/protein_id="BAB82275.1"
/db_xref="GI:5381178"
/translation="MSNHSRROQKHSHHTPLRVINFLVIFILLVSVSLFMYRHHF
LAPHLAVIYGVVLVJLILASLELCINKKARIEFTIILVLASIFVATLGYFKSTIDL
TNLNKATASEISEMSVIVPKDSKINTINEAVSKLAPVKNDTSNITDLIEHKSEGI
SITPKQDTSODAYNRINKGDSQAMVLLNAYVSLIELSTPDFKSOIKTIYIKIKKI
NRKNTHEGVNIYISGIDTFGSIIVSRSDVNIIMTVNTNTHKVLTTTPRDAYK
IPDGGHQYDKLTHAGLYGVETSMKLENLDYIARINFSSFLKLDLGGVTV
YNDQAFPSKHGDFPVQVTLNSEQALGFVRERYSLQGGDNDGRNOEKVIAIINK
LASQSVTKLNSITSOLOQTSVQTNMTIDNLDLNNLSTGQRTFVESQALTGHGVTG
ELPSYAMPQALNMSIDQSSLSNAKSKIKNTHEE"
  2975..3706
  /gene="cpsiaa"
  2975..3706
  /gene="cpsiaa"
/codon_start=1
/transl_table=11
/product="cpsiaa"
/protein_id="BAB82276.1"
/db_xref="GI:5381179"
/translation="MIDIHSHIVFDVDDGPKTLEESLSLIEESYRGCVRIIVTSRHR
KGMFETPEDIIFKNFSIVKHEAKRFEHLLOILYGGELYTSDMLEKLLKLOIPTLNT
KFLALIESMOTSKWDHTALSNLMLGITPVVAHIERYNALENOKERKVELINMGCVT
QINSHILKOKLFDKHKRFRKRYFLEENLVHFVASDMHNDLVRPPLAEAYKIC
RDFGKERANQLFIENASILKNHYI"
  3715..4407
  /gene="cpsiaa"
  3715..4407
  /gene="cpsiaa"

```

```

/function="chain length determination (putative)"
/codon_start=1
/transl_table=11
/product="cpsiaa"
/protein_id="BAA82277.1"
/db_xref="GI:5381180"
/translation="MNKIANTVEINIFNLLKLLKKKLEITFVAIAPATAGLFTYSLF
ITYPTQISSITRIYVINPNNISITAOQLQAGSFANDYKEITITSDVLEKVISSEKL
NYPSSQLQKITVSLKIDTRVISISVEDGANPRMSQRLANSVREAAVSKAVTOVEDI
TTLEKGNLPKAPSSPNIKKNVLIGFIVGAGLSTIVLVIIMGILDDRVTBEDIEKALGL
TSLGIVPDLNKL"
  4418..14922
  /note="corresponds to DDBJ Acc#:AB017355"
  4418..5107
  /gene="cpsiaa"
  4418..5107
  /gene="cpsiaa"
/function="chain length determination (putative)"
/codon_start=1
/transl_table=11
/product="cpsiaa"
/protein_id="BAA82278.1"
/db_xref="GI:5381181"
/translation="MTRLEIVDSKLPQAKKTRFYFNAJRTNIQFSGENKILAITSVR
EKGKSTTSLSLSLAQAQFKTLLIDATRNSVMSGTFKATGTIKGLTNYLSGNADL
GDICETNPRLMVPSGKVPNPPTALLQNAFYNNKIEAKNIFDIITDTPPIGLV
DAAITANACDGFILVTQAGRIKRNYYEKAKEQMSGKFLGIILNKVSESATYGDY
GNYGKRDRKK"
  5243..5248
  /evidence=experimental
-35_signal
  5265..5270
  /evidence=experimental
-10_signal
  5334..6509
  /gene="cpsiaa"
  5334..6509
  /gene="cpsiaa"
/function="glucosyltransferase"
/codon_start=1
/transl_table=11
/product="cpsiaa"
/protein_id="BAA82279.1"
/db_xref="GI:5381182"
/translation="MVLKYSFYIIFSSSLFFIFKNSFTTTLSPFSFIFAMNSILYL
LNSFLKYRYKYSKFSRDTKVLITNKSLSKMTFRNKYDHNHYIACVILDSSEKDCY
DLKHSRLIINKDALTSCLTVDQAFINIPTELFKYQIOIDIINDIEMGVIVVNV
VEALSPDNIGEKRIOTFEYSVITYSMKFKYSHLIAKRFELDTITGAIIGLICGIVAI
FLYQIRKGGGPAIFSONRVGNRGRIFRFKPSRMVDAEQIKKDLLVHNMOTGLMPK
LDDPRITIKGIRKTSIDELPQFYNVLKGDMSLVGTPTPTVDETERNSTOKRRLS
FKPGITGLMOISGRNNITDFDEIVLKDVOYINEMSWISDIKILLTLTKVVLGTTGAK"
  6533..6982
  /gene="cpsiaa"
  6533..6982
  /gene="cpsiaa"
/codon_start=1
/transl_table=11
/product="cpsiaa"
/protein_id="BAA82280.1"
/db_xref="GI:5381183"
/translation="MKICLVSGGGHLLHNLKLPFGKNKDRFWTTFDKEDARSILRE
EIVYHCFPTNRRNVLNNTILAFKVLKRKRPDVIISSGAAVVPFFYIGLFGCKT
VIIEVDFDRIDKPTLTCKLVYPVTDFKFIQVOMEEMKKVYPKAINLGGIF"
  6982..7524
  /gene="cpsiaa"
  6982..7524
  /gene="cpsiaa"
/function="galactosyltransferase"
/codon_start=1
/transl_table=11
/product="cpsiaa"
/protein_id="BAA82281.1"
/db_xref="GI:5381184"
/translation="MIFVTVGTHQOQFNRLIKEVDRLKGTGATDQEVFQTGYSDPEP

```



```

/transl_table=11
/product="putative transcriptional regulator"
/protein_id="AAK33815.1"
/db_xref="GI:13622062"
/translation="MR100LHYIKIVCEGSMNAKQLFITOPSLSNVAVKLEMEWG
ITIFNRNPKGITLTKDGEVLEFVAROIIETOSLLEDRYKNNHTGRELFSVSSOHYAFV
VNAFYSLLKRTDTRVELFRETRETHWELIDYKFRSEIGVLFINDNRDVLTKLFD
NHLTSLFRAQPHIFVSKNSNPLRATLSLMDLDRFPYLSYDOGHNSFFSEEMMS
KHEKANLSKVRGRTFLNLMIGLDGYTVASGILNSLNGDOIIVAILDVPDEIDVFI
3062. .3598
/gene="SPY0899"
CDS
3062. 3598
/gene="SPY0899"
/EC_number="2.4.2.10"
/translation="Best Blastp hit = dbj|BAA02291.1| (AB028896) ORF1
[Streptococcus agalactiae]"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK33816.1"
/db_xref="GI:13622063"
/translation="MNVTEQNSHQILIOKLVSHVYTLFRDELKLVERTPSPILGGEF
PAHLVQSELDTIVAAIDTLMQORLIESTFWESEAFKLNKTKLDIVDNWIKGVDLHI
DLQSQSEVFOIIGDKRIRVFGLSDVFSSLYSALSKEAPIPDVLYEHIQIRVNLEE
BATIKYQSKVPTGEFE"
3875. .4567
/gene="pyrF"
CDS
3875. 4567
/gene="pyrF"
/EC_number="4.1.1.23"
/translation="Best Blastp hit = sp|Q9ZHA7|DCOP_STRPN OROTIDINE
5'-PHOSPHATE DECARBOXYLASE (OMP DECARBOXYLASE) (OMPDCase)
>gl|4009483|gb|AAC95452.1| (AF068902)
orotidine-5'-decarboxylase PyrF [Streptococcus
pneumoniae]"
/codon_start=1
/transl_table=11
/product="putative orotidine-5'-decarboxylase PyrF"
/protein_id="AAK33817.1"
/db_xref="GI:13622064"
/translation="MKEERPIALDFEETKAFILDFPAEEKLVYKIGMELYAQQ
PDIVRYIKSLCHNVFLDLKHLIDPNTVRAMAVYKELDIDMATVHAAGGVEMLKAARE
GLGQPTLIATVLTSTSDQMDGQNIQTSLLESVLHYSKGAAKAQLDGAVCASQEV
EAKVATPTGTCLTGPIRGKNSIGDKRVMPNQRIRIGSDYIYVGRPIQAKDPV
AAQVATKAERAG"
4625. .5254
/gene="pyrE"
CDS
4625. 5254
/gene="pyrE"
/EC_number="2.4.2.10"
/translation="Best Blastp hit = gb|AAC95453.1| (AF068902) orotate
phosphoribosyltransferase PyrE [Streptococcus pneumoniae]"
/codon_start=1
/transl_table=11
/product="putative orotate phosphoribosyltransferase"
/protein_id="AAK33818.1"
/db_xref="GI:13622065"
/translation="MTLASOIAITQLLDIKAVYLPEDPTWASCIKSPYTDNRVTL
YKTRDLIENGFEVETIKAHPEVEVIAGTAGIPGAILIADKMTLFPFVIRSKPDH
GAGNIGRVLKQKQMVIIEDLISTGVSVDAAAASREGADVLGVVVAIFYELPKAS
QNFKEAGIKLITLSNYTELIAVAKLAGYIITNDGLHLHLKFKEDQVNNQQ"
5451. .6905
/gene="amiC"
CDS
5451. 6905
/gene="amiC"
/translation="Best Blastp hit = pir|IC75543
6-aminohexanoate-cyclic-dimer hydrolase - Deinococcus
radiodurans (strain R1)
>gl|6457907|gb|AAF09821.1|AE001885_5 (AE001885)
6-aminohexanoate-cyclic-dimer hydrolase [Deinococcus

```

```

radiodurans]"
/codon_start=1
/transl_table=11
/product="putative amidase"
/protein_id="AAK33819.1"
/db_xref="GI:13622066"
/translation="MTYQDATAMIAVQTGOTTPLELVTOAIYKAKKLNPTLNAISE
RFEALAEAKQRDFSGLPFAGVPLFLKDLGOELGHSSTSGSLFKEYQATKTLDFVK
RELAAGFIILGRSNTPEFGFKNISDLHGPNVLRNDRNAGSGGAAALVSSGIS
ALATAGSGGSGIRIPASFNGLIKLPSRGPMPVPGSYRSMOGASVHFALTTKSVOTR
NLVLYLQMEQMESPPFLATLTKDSIYQSLQRPILTIAFYORLSDGPSVSLDTAKALQA
VWLRQGHQVLEEEFEPVNMTEVIRHYIIMNSVETAMPADIEDFGRPMTKDDHET
MTWAIYQSGKIDIPAWRISQVLQRWDIYSATMFGSLNYPNTALPNLTGCPAISLPYETREG
DSKLMANLAAQAEIIFSEEQFNLFVFGKSLAINPTALPNLTGCPAISLPYETREG
LSNGIQLIAAKREDLLGIAEQFEAAGLLKIPQ"
6987. .7886
/gene="SPY0903"
CDS
6987. 7886
/gene="SPY0903"
/translation="Best Blastp hit = sp|P42400|YCKB_BACSU PROBABLE ABC
TRANSPORTER EXTRACELLULAR BINDING PROTEIN YCKB PRECURSOR
(ORF2) >gl|7442312|pir|ID69760 amino acid ABC transporter
(binding protein) homolog yckB - Bacillus subtilis
>gl|2632624|emb|CAB12132.1| (Z99105) similar to amino acid
ABC transporter (binding protein) [Bacillus subtilis]"
/codon_start=1
/transl_table=11
/product="putative ABC transporter (binding protein)"
/protein_id="AAK33820.1"
/db_xref="GI:13622067"
/translation="MLYYSYLNKGVTKKSVTLISGLASLLAACAPHQSKSSWDI
IKEKGLVKVATPTGYQYTSFYNNDELGVYEYDMVKEIKRLNIKYKVFQDQAF
SVDSGSTEYKVAQKGAELTYDNTGVDYLVNDVANGRTDFIPNDYPAOKLFEVDYM
LSONPMLNKMVSVOYNPTQEGIVMKNKDDSLKKIKDAYIKDKIKGSLKIKSETIYA
GQDLTPFKDKKIPVIDIKDNV"
7915. .8565
/gene="SPY0904"
CDS
7915. 8565
/gene="SPY0904"
/translation="Best Blastp hit = sp|P42399|YCKA_BACSU PROBABLE
AMINO-ACID ABC TRANSPORTER PERMEASE PROTEIN YCKA
>gl|7442388|pir|IC69760 amino acid ABC transporter
(permease) homolog yckA - Bacillus subtilis
>gl|710628|dbj|BAA06425.1| (D30762) homologue of glutamine
permease of H. influenzae [Bacillus subtilis]
>gl|1805409|dbj|BAA08971.1| (D50453) homologue of
glutamine permease of H. influenzae [Bacillus subtilis]
>gl|2632623|emb|CAB12131.1| (Z99105) similar to amino acid

```

Query Match

Best Local Similarity

Matches

413; Conservative

0; Mismatches

229; Indels

0; Gaps

0;

QY

1

ATGATTACATAGTTTATTAACTAGCTATCTGCTGGTTCGATTCCATCTGCTC

60

|||||

|||||

|||||

|||||

|||||

Db

11450

ATGAAATTAATCTTTTATTACCATTCGCTTACTTGGTCTTATCCAAAGTGACTA

11391

QY

61

TGGATTGACAAAGTATCTTTTAAATCAATCTAGCGCAGCATGGTCTCGTACACCTGA

120

|||||

|||||

|||||

|||||

|||||

Db

11390

TGGATTGACAGTACTTTTACCACATCACTTACGAGAGCATGGATCAGGAAATGACTGA

11331

QY

121

ACGACCAACACCTTCCCATTTTAGGTAAAGAACTGGTATGGCAACCTTTTGTGATTGAC

180

|||||

|||||

|||||

|||||

|||||

Db

11330

ACCAACAATCTTTTCGGATTTTAGGTGTCAAGCAGGACACGCTACTTACGTATTGAT

11271

QY

181

TTTTTCAAGGAACCCCTAGCAACGCTCTTCCGATTATTTTTCATCTACAGGCGTTCT

240

|||||

|||||

|||||

|||||

|||||

Db

11270

ATGTTTAAAGGACACTTTTCAATATTTTACCAATATTTTGTGATGACTTCAATTCA

11211

QY

241

CCCTCATCTTTTGACATTTTGGCTGTTATCGGCATACCTTCCCTATCTTTTGCAGGATT

300

|||||

|||||

|||||

|||||

|||||

Db

11210

TCCATTCTGCTATCGGCTTTTTCGGAGTTTATAGGGCATACTTTTCTCTTTTGGCAACTTT

11151

```

/product="putative purine nucleoside phosphorylase"
/db_xref="GI:19748121"
/translation="MSIHIISAKKGGDIADKILLPCQDPAKRFIAENFLEDAVCFNVRN
MFGYTGTYKGRVSMVCTGCMGPSIYARELIVDYGKTLIRVGTAGADPEVHVRN
LVLAQAATNSNIIRNDPEPFDQIADFLGDKAYHIAEMGVTHVGNVLSDFY
TNPERNNMKGKGVKAEENAAALYTLAAQHVKAIGIMTISNLDNLDPTDPTTAEER
QTTFDMMKVGLETLIAND"
1062..1856
/genes="spym18_0956"
1062..1856
/genes="spym18_0956"
/note="best blastp match gblAAK33814.1 (AE006539)
histidine protein kinase [Streptococcus pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="histidine protein kinase"
/protein_id="AAL97597.1"
/db_xref="GI:19748122"
/translation="MPMTNQTDLILDVYAYNAFRIAKALPNIKTVLYLLEMLER
REINLAFALHAENATEIDQYHCSLWLNQSDLEQYANILDELYKVRKNGAIIDF
RSVSPILYRLFRLITSEIPNFKAYIFDTKNDOYDTWHFOAMLESDEHVEFKAYLSQK
SRNVTYSKADMLTLTSLPOEIKDLVFLLRHFKAQRNPLAHLIKPDEEELHRTTHE
SSQAFLENITLATFSGVIVRRPEFYDDMMNAIKKELSLRQSIY"
complement(1936..2841)
/genes="cpsy"
/note="spym18_0957"
complement(1936..2841)
/genes="cpsy"
/note="best blastp match gblAAK33815.1 (AE006539)
putative transcriptional regulator [Streptococcus pyogenes M1
GAS]"
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator"
/protein_id="AAL97598.1"
/db_xref="GI:19748123"
/translation="MRIOQLHYIIKIVECGSMNAAKOLFITOPSLSNVAKDLEMENG
ITIFNRNPKGITLTKDQVEFLSYARQITEOTSLLEDYRKNHTGRELFVSSSHAFY
VNAFVSLKRTDMRYELFLFRETMTWETIDQVKNRSEIGVLFINDYNROVLTKLDD
NHLTASPLKQAQPHIFYSKSNPLATKSLSSMDLLRDPYLSYDQGIHNSFYSEEMKS
OMPHNSIVSDRATFLNMLGDTYVASCILNSNLNGDQIVAIPLDVPDEIDIVFI
KHEKANLSKMGFERFIYXLEEVTDF"
3063..3599
/genes="spym18_0958"
3063..3599
/genes="spym18_0958"
/note="best blastp match gblAAK33816.1 (AE006539)
conserved hypothetical protein [Streptococcus pyogenes M1
GAS]"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAL97599.1"
/db_xref="GI:19748124"
/translation="MNVTEQNSHQTILQKLLVSIHYLTFLRDELKLVERTPTSLGGE
PAHLVQSSGLDIIAADDLDQMOORLIESTFWKESAFKLMNKTLDDIDNMWIKGVDDHL
DLCSQKEVFQIIGDKRTRVFGVLSDFSSLSKVSALSKEAPIDPLVYEHIVQRVNL
EAFIKYQSPKVPPTGEF"
3876..4568
/genes="pyrf"
/note="spym18_0959"
3876..4568
/genes="pyrf"
/note="best blastp match gblAAK33817.1 (AE006539)
putative orotidine-5'-decarboxylase PyrF [Streptococcus
pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="putative orotidine-5'-decarboxylase"
/protein_id="AAL97600.1"
/db_xref="GI:19748125"
/translation="MKERETIALLDFSSFEETKAPFLDLFPAEEKLYVKIGHELYQAQ

```

PDIVRYIKSLGHNVFLDLKHDIPNTVRAMAVLKELDDIMATVHAAGGVEMLKAARE
GLQDPTLIATVLTSTSEDDMGDQNIOTSLLESVHLVSKGAQAQLQDGVVCSAQEV
EAKAVPTGFTCLTPGIRPKGSGNIGDQKRVMTPNQARRIESDIIVVGRPTTQAKDPV
AAYOAIKAEWAG*
4626..5255
/gene="pyrE"
/note="spym18_0960"
4626..5255
/gene="pyrE"
/note="pyrE"
/note="best blastp match gb|AAK33818.1| (AE006539)
putative orotate phosphoribosyltransferase [Streptococcus
pyogenes M1 GAS]
/codon_start=1
/transl_table=11
/product="putative orotate phosphoribosyltransferase"
/protein_id="AAL97601.1"
/db_xref="GI:19748126"
/translation="MTLASQIATQLLDIRAVLYLKPEDPFTWASGIKSPYVTONRVTL
YKTRDLIENGFEVETIRAHPEVEVIAGTATAGIPIHGAIIDAKMTLPFAVIRSKPKDH
GACNOIEGRVLKQKMWIILEDLISGTGSVLDRAAASREGADVGLVVAIYFTFLPKAS
ONKEAGIKLITUSNTELIARVAKLOGIITNDGLHLLKFKEDQVNMQQ"
5452..6906
/gene="amic"
/note="spym18_0961"
5452..6906
/gene="amic"
/note="best blastp match gb|AAK33819.1| (AE006539)
putative amidase [Streptococcus pyogenes M1 GAS]
/codon_start=1
/transl_table=11
/product="putative amidase"
/protein_id="AAL97602.1"
/db_xref="GI:19748127"
/translation="MTYQDATAMAIQVOTGQTPPLELVTAIYKAKKLNPLNATSE
RFEALAEAKRDFSGLPFAGVPFIKLDGLOELKHSSTSGSRLFKFYONATKIDLFVK
LEALGFIILGRQNTPEFGKNIIDSLHGPVNLKPDNTRNAGSSGSAALVSSGIS
ALAPASDGGGSRIPASFNGLIGKPSGRMPVPGSYRSGWOGASVHFALTSVRDTR
NLIYIQHOMESFPFLAVLTKDSIQYQRPULTAFYORLSDGSPVLDLAKAKQA
VWLRGQHOLVEEFPVNMTEVIRHYIYMSVETAMFADIEDTLGRPMKDDMET
MTWAIYOSGKIDIPAWRYSOVLQKWDYTSATMASPHETYNLLLTFTTNPAPKHGLVP
DKLMAANLAQAEIIFSEEQFNLVETMFGKSLAINPYTALPNLTGPAISLPTYETKEG
LPMGIOLIAAKGREDDLGLIAEQFEAAGLLKIPQ"
7024..7887
/gene="spym18_0962"
7024..7887
/gene="spym18_0962"
/note="best blastp match gb|AAK33820.1| (AE006539)
putative ABC transporter (binding protein) [Streptococcus
pyogenes M1 GAS]
/codon_start=1
/transl_table=11
/product="putative ABC transporter (binding protein)"
/protein_id="AAL97603.1"
/db_xref="GI:19748128"
/translation="MKKSVTLLSGLASLLLAACAPHKSQKSSWNTIKKGVKLVKATP
GTYPQTSFNDNNELCYEVDWKEIGKRLNKKVFETGFDQAFVSDSGRVDISLN
NFDITPARKYINISIPYIGYGMIVRADGSSNMAKDLSDMKGKAAGASGETYMK
VAQQAEMVYDNTVDYLVNDVANGRTDFIPNDYPAQKLFVDYMLSQNPRLNVKMS
DVQNPTEQIGVNNKDDSLKKIDAVIKDMKKDGLSKKISETIYVAGQDLTKPFGKDK
KIPVDTKDVN"
7916..8566
/gene="spym18_0963"
7916..8566
/gene="spym18_0963"
/note="best blastp match gb|AAK33821.1| (AE006539)
putative ABC transporter (permease) [Streptococcus
pyogenes M1 GAS]
/codon_start=1
/transl_table=11
/product="putative ABC transporter (permease)"
/protein_id="AAL97604.1"
/db_xref="GI:19748129"
/translation="MINIPLMDSGLGVLGSLPYTLIGISULLSLTGLVFLGLLALLGR

SRQPLHLVLRVYISIMRGVPMIVLVFLVFLPYGLPALLCAYLGFNSVSAAYI
SEVFRSISAIIDKQWEAAKALGLPILMVAIYLPQAFRIAPVPPGNVIIDMWKSSS
LAAMITVPDIFQNAKIIGRENDYMSYILVAFIYLAFLRLRYOEFLNKLALV*
8699..9352
/gene="ung"
/note="spym18_0964"
8699..9352
/gene="ung"
/note="best blastp match gb|AAK33822.1| (AE006539)
Query Match 42.9%; Score 275.6; DB 1; Length 11573;
Best Local Similarity 64.3%; Pred. No. 7.8e-59;
Matches 413; Conservative 0; Mismatches 229; Indels 0; Gaps 0;
QY 1 ATGATTACAATAGTTTATTAATCCTAGCTATCTGCTGGGTTCGATTCATCCTGGTCTC 60
DB 11478 ATGAATTTACTACTTTTATTAACCAATTCCTATTTACTTGGTCTTATTCACATGGA 11419
QY 61 TGGATTGGACAAGTATTTTCAATCAATCTACCGCAGCATGGTCTGCTAATACATGGA 120
DB 11418 TGGATTGGACAGTACTTTTACCACATCACTTACGAGCATGGATCAGGAAATACATGGA 11359
QY 121 ACGACCAACACCTTCGCGCTTTTAGGTAAAGAACTGGTATGCAACCTTTGTGATTGAC 180
DB 11358 ACCCAATATCTTTTCGGAATTTTAGGTGTCAGGCAAGCACTACCTTAGCTATTGAT 11299
QY 181 TTTTCAAGGAACCTTAGCAACGCTCTTCCGATTATTTTTCATCTACAAAGCGTTTCT 240
DB 11298 ATGTTAAAGGGACACTTTCAATATGCTTACCAATATTTTGGTATGACTTCAATTTCA 11239
QY 241 CCTCTCATCTTTGGACTTTTGGCTGTTATCGGCCATACCTTCCTCTATCTTTGCAAGATT 300
DB 11238 TCCATTGCTATCGGCTTTTTCGCGAGTTTTCAGGCATACCTTTCTTATTTTGGCACTTT 11179
QY 301 AAAGGTGGTAAGGCTGTCGCAACAGTCTGCGAGTATTTTTCGATTTTCGCGCTATCTTC 360
DB 11178 AAAGGTGGTAAGGCTGTCGCAACAGTCTGCGAGTATTTTTCGATTTTCGCGCTATCTTC 11119
QY 361 TGTCTCTACCTTCGATTATCTTTTGGAGCTCTCTATCTTGGCAGTATGATTTTCACTG 420
DB 11118 CTCTTTTTCATCATCTATCTTTGTTTATTTAGTTCTCTATTTATTTAGTATGATATCTTTA 11059
QY 421 TCTAGTGTACACATCGATTGCGGCTGTTATCGGGGTCTGCTCTTTTCCACTTTTGGT 480
DB 11058 GCTAGTGTGCTTTCAGCTATCTGTTGGTGTATCTGTTTAAACATTTTCTGCGCATTCAT 10999
QY 481 TTTTCTCTAGTAACTATGATCTCTCTCTTCATCGCTATTTATCTTAGCAGTCTGCTAGTTG 540
DB 10998 TTTCTTTTACCAATATGACTACTTTTAACTTTTATGATATTTACTTTCGCTTCATT 10939
QY 541 ATTATCATTCGTCATAGGACAAATATAGCTCGTATCAAAAATGAAATGAAATTTGGTC 600
DB 10938 ATTATTAAGACACAAAGATAACATTAGTCGTATCAACATCATCATGAAATCTAATA 10879
QY 601 CCTTGGGGATTGAACCTTAACCCATCAAGATCCTTAAATAATAA 642
DB 10878 CCTTGGGGCTAAATTTAAAGCAACAAGTCCCTTAAATAATAA 10837
RESULT 9
AE014149/c 5334 bp DNA linear BCT 19-JUL-2002
LOCUS Streptococcus pyogenes MGAS315, section 14 of 37 of the complete
DEFINITION genome.
ACCESSION AE014149 AE014074
VERSION AE014149.1 GI:21904329
KEYWORDS
SOURCE Streptococcus pyogenes MGAS315.
ORGANISM Streptococcus pyogenes MGAS315
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
REFERENCE 1 (bases 1 to 53354)
AUTHORS Beres, S.B., Sylva, G.L., Barbican, K.D., Lei, B., Hoff, J.S.,

TITLE	Mammarella, N.D., Liu, M.-Y., Smoot, J.C., Porcella, S.F., Parkins, L.D., McCormick, J.K., Leung, D.Y.M., Schlievert, P.M. and Musser, J.M.	CDS	/gene="clpX" /note="Spym3_0604" 1883..3112 /gene="clpX" /note="best blastp hit: gb AAK33805.1 (AE006538) putative ATP-dependent Clp protease subunit x [Streptococcus pyogenes M1 GAS]" /codon_start=1 /transl_table=11 /product="putative ATP-dependent Clp protease subunit x" /protein_id="AAW79211.1" /db_xref="GI:21904331"
	Genome sequence of a serotype M3 strain of group A Streptococcus: Phage-encoded toxins, the high-virulence phenotype, and clone emergence		
JOURNAL REFERENCE AUTHORS	Proc. Natl. Acad. Sci. U.S.A. 99 (15), 10078-10083 (2002) 2 (bases 1 to 53354) Beres, S.B., Sylva, G.L., Barbican, K.D., Lei, B., Hoff, J.S., Mammarella, N.D., Liu, M.-Y., Smoot, J.C., Porcella, S.F., Parkins, L.D., McCormick, J.K., Leung, D.Y.M., Schlievert, P.M. and Musser, J.M.		
TITLE	Direct Submission	gene	/gene="Spym3_0605" /note="best blastp hit: gb AAK33806.1 (AE006538) conserved hypothetical GTP-binding protein [Streptococcus pyogenes M1 GAS]" /codon_start=1 /transl_table=11 /product="putative GTP-binding protein" /protein_id="AAW79212.1" /db_xref="GI:21904334"
	Submitted (14-JUN-2002) Laboratory of Human-Bacterial Pathogenesis, Rocky Mountain Laboratories, NIAID, NIH, 903 South Fourth St., Hamilton, MT 59840, USA		
FEATURES	Location/Qualifiers	CDS	
source	1..53354		
	/organism="Streptococcus pyogenes MGAS315"		
	/strain="MGAS315"		
	/serotype="M3"		
	/db_xref="taxon:198466"		
	/note="group: A"		
	123..986		
	/gene="thya"		
	/note="Spym3_0601"		
	123..986		
	/gene="thya"		
	/note="best blastp hit: gb AAK33802.1 (AE006538) putative thymidylate synthase [Streptococcus pyogenes M1 GAS]"		
	/codon_start=1		
	/transl_table=11		
	/product="putative thymidylate synthase"		
	/protein_id="AAW79208.1"		
	/db_xref="GI:21904330"		
	/translation="MLTERLEMTKADQIFKANIQKIINEGSLSEQARPKYKDGRTAH SKYTGAFAVDLAKGEFPITLRPIKSAIKLEFWIYODQNSLDVLEAKYVNIWV NENEVDOTRIGORYGAVYKKHDIISKILKQLEPNRRNVLNDYEFETKGLL PCAFQIMFVRVGGEDYLDASLITRSDILVAHHINAMQYVALQMMIAKHPGKGLG FFYFVNNLHYDNQFOAQELLKRPVQSPKLVNLVNPRTNFFDIKPDPDFELONYDP VKPOLHFDLAI"		
	1066..1563		
	/gene="dyr"		
	/note="Spym3_0602"		
	1066..1563		
	/gene="dyr"		
	/note="best blastp hit: gb AAK33803.1 (AE006538) putative dihydrofolate reductase [Streptococcus pyogenes M1 GAS]"		
	/codon_start=1		
	/transl_table=11		
	/product="putative dihydrofolate reductase"		
	/protein_id="AAW79209.1"		
	/db_xref="GI:21904331"		
	/translation="MTKEIIAIWADEAGLIGVAGKLPWLPKLEHFKKTTLHQAIL MGRVTFEGMCKRLPORQTLVMTNRNDYQVDEVLTMTSIEKVLWEVHAQDKTLYIIGG NKVLEAFNGYFDRIKTVIHHRFGDTYRPNLDFSHFTQESQTFYARDAKNPYDFVT ALXHK"		
	1583..1753		
	/gene="Spym3_0603"		
	1583..1753		
	/gene="Spym3_0603"		
	/note="best non-GAS blastp hit: gb AAK75657.1 (AE007451) hypothetical protein [Streptococcus pneumoniae TIGR4], and gb AAL00232.1 (AE008512) hypothetical protein [Streptococcus pneumoniae R6]"		
	/codon_start=1		
	/transl_table=11		
	/product="conserved hypothetical protein"		
	/protein_id="AAW79210.1"		
	/db_xref="GI:21904332"		
	/translation="MQRSEIFGVFTAFVLCVLCVCAIPAFKKNRYGLASFLNFAFTNL VNSTHAFYGTLF"		
	1883..3112		
	/gene="clpX"		
	/note="best blastp hit: gb AAK33808.1 (AE006538) putative ATP-dependent Clp proteinase (ATP-binding subunit) [Streptococcus pyogenes M1 GAS]"		
	/codon_start=1		
	/transl_table=11		
	/product="putative ATP-dependent Clp proteinase (ATP-binding subunit)"		
	/protein_id="AAW79214.1"		
	/db_xref="GI:21904336"		
	/translation="MTHFSKDPFVNMDDIFNQLMANMGYRSENPRLVNGREITPE EFQHYRQQLPVAATKATNSQMLTPKADSVLTQLGTNLTOEARQHLDPVIGRNKEI QDTAEILARRTKNNPVLVDAGVGTAVTEGLAQAVTNGDVPAAIKNKKEIVSDISL EAGTQVRGSEETIQNLIOEVEAGNIILFFDEIQIVGAGATSSDSGSKGLILKPK ALSRGELTIGATTODEYRNTILKNAALARRNEVKNAPSAEDTFHILMGINLVEQ HHHTITPDNLVKAADVSIQYIPQBSLPDKADLDLDMTAHLAAQHPVDTLKTLEI AKQESQEKAVAKEDFEKAAKTRITETLQKIQEQHNQSNQVTTATVNDIAESVERLTG		

```

IPVSNMGTNDLRLXGISRLKSHVIGQDEAAVAARAI RRNRAGFDDCKRPIGSEFL
VQPTGVGKTELAKQIALDLFGSKDAIIRLUMSEYNDRTAVSKLIGTAGYGVYDDNN
TUTYRPNRYAIVLLDEKAPQIITLLQVLDREKRLTDGQNTINFKNTVIATTS
NAGFGQODTETSESIMDRAPYFPEFLNFRNSIIKFNHLOKESLEEIVDLMLAEVN
QTTAKGISLTTTDDAKAHLIDLGYNHANGARPLRIIEQEI RDRITDYLDHPVEVK
LOALLKEGOLVIRQNDQ"
7147..7830
/gene="rpiA"
/gene="rpiA"
/notes="Spym3_0608"
7147..7830
/gene="rpiA"
/notes="best blastp hit: gb|AAK33809.1| (AE006538) putative
ribose 5-phosphate isomerase [Streptococcus pyogenes M1
GAS]".
/codon_start=1
/transl_table=11
/product="putative ribose 5-phosphate isomerase"
/protein_id="AAW79215.1"
/db_xref="GI:21904337"
/translation="MEALKAKIAGVTAAYVTDGMTIGLGTSTAYFYVEEIGRRVKQE
GKQVGVTSSTVYVQDASKMVEHLPLKSIDIDSLTVDGADVDKRNFKNGIKGGGAALL
MKIIVATPPTSVYIWWVDASKMVEHLCAFKLPVEVVOYGADRLFRFVEKAGYKPSPMKK
GDSRLVTDQMNYIIDLDLGCIDPVAFCFHLLDGTGVGVHEGLFNGMVDKVIASKDG
TVLEAPTAG"
7907..9118
/gene="deob"
/notes="deob"
/notes="Spym3_0609"
7907..9118
/gene="deob"
/notes="best blastp hit: gb|AAK33810.1| (AE006538) putative
phosphopentomutase [Streptococcus pyogenes M1 GAS]".
/codon_start=1
/transl_table=11
/product="putative phosphopentomutase"
/protein_id="AAW79216.1"
/db_xref="GI:21904338"
/translation="MSKFNRIHLVLDSVIGI GAAPDADKFFNAGVADTSDTLGHISE
AAGLSVPNNKIGLGNISRPILKTVPTEDNPTGYVKLEEVSLGDTKPTGHWHEIMGL
NITEPDTFWNAGPEBILPKIEEFGSRKRIIRKANKPYSGTAVIDDGFGQMTEGELIV
YSADPVLQIAHEDIIPVEELYKICEYARSITFEPALLGRIIARPYGPDGPNFTRT
ANRHDYAVSFODTIVPKLADAGVPTYAVGKINDIFNGSGITNDMCHNKSNSHGIDTL
IKTLQLPEFTKGFSTNLVDFDANFCHRRDPEGYRCDLHEFDNRLPEI IANKKEDDLL

Query Match      42.7%; Score 274; DB 1; Length 53354;
Best Local Similarity 64.2%; Pred. No. 1.8e-58;
Matches 412; Conservative 0; Mismatches 230; Indels 0; Gaps 0;

Qy 1  ATGATTACATAGTTTTATTAACTTAGCCTATCTGCTGGTTCGATTCCATCTGGTCTC 60
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 22122  ATGAATTTACTACTTTTTATTACCATTGCCTATTATTAGTTGGTCTTATCCAACTGGACTA 22063

Qy 61  TGGATTGGACAAGTATTCTTTTCAATCAATCTACGCCAGCATGTTCTCTGCTAACTGGA 120
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 22062  TGGATTGGACAGTACTTTTACCACATCACTTTACGAGAGCATGGATCAGGAATACTGGA 22003

Qy 121  ACGACCAACACCTTCGCCATTTTATAGTAAGAAAGCTGGTATGCCAACCTTTGTGATTGAC 180
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 22002  ACCACAAATACTTTTTCGGATTTTAGGTGTCAAGGCAGACACGCTACCTTAGCTATTGAT 21943

Qy 181  TTTTTCAAAGCAACCCCTAGCAACGCTGCTTCGGATTATTTTTCATCTACAAGGGGTTCT 240
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 21942  ATGTTTAAAGGGACACTTTTCAATATGTTTACCATAATTATTTTGGTATGACTTCAATTTCA 21883

Qy 241  CCTCTCATCTTTGGACTTTTGGGTGTTTATCGGCATACCTTTCCCTATCTTTGCGAGGATT 300
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 21882  TCCATTGCTATCGGCTTTTTCGCAGTTTTTAGGCGCATACTTTTCCATTTTTTGGCAACTTT 21823

Qy 301  AAAGGTGGTAAGGCTGTGCAACACAGTCTGGAGTGAATTTTCGGATTTGCGCCTATCTTC 360
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 21822  AAAGGTGGTAAGGCGGTAGCAACAAGTCTGTTGCTAGGCTTTTGTCTCCGTTATAT 21763

Qy 361  TGCTCTACCTTGGCATATATCTCTTTTGGAGCTCTCTATCTTGGCAGCTATGATTTCAC 420
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

Db	21762	CTCTTTTATAGCATCTATCTTTGTTTATAGTCTCTATTTATTTAGCATGATATCTTTA	21703
Qy	421	TCAGTGTACAGCATCGATTGGCGCTGTGTATCGGGGTTCTGCTCTTTTCCACATTTTTGGT	480
Db	21702	GCTAGTGTGGTTTCAGCATCGTTGGTGTGTATCTCTTTTAACATTTTCTCGCCATTCAT	21643
Qy	481	TTTATCTCGAGTAAGTACTCTCTTCATCGCTATATCTTTAGCATGCTGTAGTTTG	540
Db	21642	TTTCTTTTACCAATATGACTACTCTTTTAACTTTTATGATATATCTGCTGCTTCATT	21583
Qy	541	ATATATCATTCGTCTATAGGACAATATAGCTGCTGATAAATAAATAAATTAACATTTGGTC	600
Db	21582	ATTATTAAGACACAAGATAACATAGCTGATCAAAACATCATCAATAAATCTAATA	21523
Qy	601	CCTTGGGATGTAACCTAAACCATCAAGATCCTAAAAATAA	642
Db	21522	CCTTGGGCTAAATTTAAGCAACAAGTGCCTAAAAATAA	21481
RESULT 10	AE006332/c		
LOCUS		13211 bp DNA linear BCT 14-MAY-2001	
DEFINITION		Lactococcus lactis subsp. lactis ILL403 section 94 of 218 of the complete genome.	
ACCESSION	AE006332	AE005176	
VERSION	AE006332.1	GI:12723916	
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
source			
gene			
CDS			
gene			
CDS			

```

/codon_start=1
/transl_table=11
/product="ribonucleoside-diphosphate reductase alpha chain
(EC 1.1.7.4.1)"
/protein_id="AAK05073.1"
/db_xref="GI:12723918"
/transl_table=11
/translation="MSLKQKDVITYFKLNNEIPVNGSIPLAKDKEAITAFFKENVE
PNRYEADTYIDKLNWLKVELEYLAEFLEKYKPEFITNLRKELSKFEFFHDFSPMAVKF
NQYAKNNGDSLYLEDFEDRVLNMLYADGELKANNLAFAMINRYPOTPTEFLN
AGRRRGEVSCFLQLQTDHMSINTINSALQSLKNGGVLNLRRAAGAPIKGI
DGAAGVVPVWKLIFEDAFSYANQIGORAGVYVLSLFHPDIMEFLSTKKENADEKIR
KYSINARELEQELSKLQOESGYPYIVNMDTVANOANPDGVISMSNLCSEILOVQTPS
VLNDQTYKVVGDTVACNLGTSNVNLMNTSSNEEGTSIEAMVRLAFISETSLDTPV
TVRGKNDMAHGLGAMGLHSFLAQNQIHYDSKEAVEFTSYFMLNLYYTLVLSNKLK
VEKSGSKNFKESAYADGYSYDFKYLTHDYFAAVSKVQELFKGIAPVTIEDWLTALRDT
VYTSGLVNSYMAVAPNGSISYINDCSSSIHPVNNRIEERQEKVKGIYYPAAGLATE
TLPYKAYOTDMRAVDYVAAATEHVDQGLSLTFLMRSTLPEGLYEMKVANNKMTTR
DLSLRNYAYKGIKIYIIRTFTDQOEEVGANQCESCVI"
complement(3713..4135)
/gene="ndrI"
/notes="L4702"
complement(3713..4135)
/gene="ndrI"
/notes="EVIDENCE BY HOMOLOG Y BIO06.05 ENERGY METABOLISM
Electron transport. BELONGS TO THE NRDI FAMILY"
/codon_start=1
/transl_table=11
/product="ribonucleotide reductase"
/protein_id="AAK05074.1"
/db_xref="GI:12723919"
/translation="MKLAFESVTGQTRRFVSKTNLNVLSIPDDIEMNEPFLITPS
YAESPTVKSIDVMDPVDFMAYNONYKHLGIGTGGRNFAGIIFTAKELSAKIQ
IPLYDFENGTPADVAAEKELATQLDKGAKVTFKNPL"
complement(4137..4313)
/gene="ndrH"
/notes="L5126"
complement(4137..4313)
/gene="ndrH"
/notes="EVIDENCE BY HOMOLOG Y BIO06.05 ENERGY METABOLISM
Electron transport. ELECTRON TRANSPORT SYSTEM FOR
RIBONUCLEOTIDE REDUCTASE SYSTEM NRDEF"
/codon_start=1
/transl_table=11
/product="glutaredoxin-like protein Nrdh"
/protein_id="AAK05075.1"
/db_xref="GI:12723920"
/translation="MVKKLSLSEHIAFDEINIDQEPFEKVIEMGFRAAPVITKDDF
AFSGFRPSELAKLA"
complement(4529..5170)
/gene="ykcC"
/notes="L5517"
complement(4529..5170)
/gene="ykcC"
/notes="EVIDENCE BY HOMOLOG Y BIO15.02 HYPOTHETICAL.
Conserved. integral membrane protein"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK05076.1"
/db_xref="GI:12723921"
/translation="MLTIILLIASYLLGAIPFGLWGKIGFKFNLDYSGNGTGTN
TFRILGVKAGISVAFDLKGLATLPLPFHNGVSPILFGLLVIGHFTFSIFDRFK
GGRAVTSAGVILGSPFLIYLLVWFIIVLWLSFMSLSVIGCAVFAIGLIIFPSI
GFLITSYDLFLFSIIIFVLAIIILRHRTNLRKIKNHCESLVPEGLNLSKOKER"
5445..7379
/gene="parE"
/notes="L0290"
5445..7379
/gene="parE"
/codon_start=1
/transl_table=11
/product="topoisomerase IV subunit B (EC 5.99.1.1-)"

```

```

/protein_id="AAK05077.1"
/db_xref="GI:12723922"
/translation="MVIDNNYDDSAIQVLEGLDANVRKRPGWYIGSTGTGLHLHWE
IVDNAVDEALSGFNGRISLVTINADGSLSVSEDEGRMPVKGHMGIPTEVIFILHAG
KFGQGGYKTSGLGHVGVSSVNALSTWLEVEITRDGTVYRORFENGHPATTLVKG
KAPKSTGTVKTEFMPDSSIFSTTEFYNIISERLNEISAFLLKEVITISLEDKRDEKEE
FHYENGQDVSYLNDKDSLTPLVYPEGQDGFVEVALQYNDGYSENLSFVNNVR
TKDGGTHEAGLKAITKAMNEYARKVNLKDEKDNLEGSYRGLTAVLSVLPEEHL
QFSGQTKDLGSLARPVDLSYSEKLSFYLMENGELAQNLRKAIKARARAAKKA
REDNRGKSKKDKLGLSLTQAQTNPNKNLILYVEGDSAGSARQGRKRFQAIL
PLRGKVINTEKAKHQDILKNEETIMTYITIGAGVADFTLADRNKYDKIIIMTDADTDG
AHQITLLTTFYRMPKPLVEAGHVIYALPPLYKMSCKGKKKEIYEAWTGELAELRK
QFQKNANLORYKGLGEMNADOLWETMDPETRTLQIQTVIDDAQAQAKRVSVLMDKVE
PRKTIENNVKFTMEETVE"
7584..7790
/gene="ykaE"
/notes="L8551"
7584..7790
/gene="ykaE"
/notes="EVIDENCE BY GENEMARK BIO16.01 Unknown"
/codon_start=1
/transl_table=11
/product="UNKNOWN PROTEIN"
/protein_id="AAK05078.1"
/db_xref="GI:12723923"
/translation="MOKSALIIDILOAFIKKLAKDAVPYGDWIKIENNRLAISAFDOK
NQPYIIVTEPKYFGNQLKFLG"
7787..8122
/gene="ykaF"
/notes="L8730"
7787..8122
/gene="ykaF"
/notes="EVIDENCE BY HOMOLOG Y BIO15.01 HYPOTHETICAL"
/codon_start=1
/transl_table=11
/product="HYPOTHETICAL PROTEIN"
/protein_id="AAK05079.1"
/db_xref="GI:12723924"
/translation="MILQDELDRNRLHQLVKYGPSAFKOSDYGLADELKNNSVSKYI
VSGVSMNSGVLTARGADIGFEAMIEDACDRKLNFQKTDEIPKFGKLIKTKE
IIEELKNV"
8115..8948
/gene="dnaQ"
/notes="L9094"
8115..8948
/gene="dnaQ"
/notes="EVIDENCE BY HOMOLOG Y BIO10.02 DNA replication,
restriction, modification, recombination, and repair. DNA
polymerase III responsible for most of the replication
synthesis in bacteria. Exhibits also 3' to 5' exonuclease
activity."
/codon_start=1
/transl_table=11
/product="DNA polymerase III, epsilon chain"
/protein_id="AAK05080.1"
/db_xref="GI:12723925"
/translation="MSEKYVVFDTTGFSAQKNEIIQIGAVKYDENNOQARFNOLY
KNORSYSTEITQTKVPADLLNFTIEILPEFLDFIKGHLVVAHNAFPDIFSFLYQ
AIIDCAITDVEAFRVDTLAESKKLLTMSYALENKQLEMDLRSHADNLCLITA
KLQYIKQIEVELSVSEVTDDELGOLDLFGDSSDEITSLRRKLNLPITKALVYQO
NLKRWREVLGISIEREYVSGSSNLRLHNSEKVKIHSDFLEKQKQANFVTIMEEE
K"
8945..9394
/gene="ykbA"
/notes="L9876"
8945..9394
/gene="ykbA"
/notes="EVIDENCE BY HOMOLOG Y BIO15.01 HYPOTHETICAL.General"
/codon_start=1
/transl_table=11
/product="HYPOTHETICAL PROTEIN"
/protein_id="AAK05081.1"
/db_xref="GI:12723926"

```

/translation="MTAIWKEKSFNQLTLDYVOLLRLTAVFVVEONCPVQELDKGD
QVAKHLWENEGOLMALCRLLPKGISFAEASIGRVLVNPEFRQNLGRGLIEKALEV

Query Match 38.4%; Score 246.6; DB 1; Length 13211;
Best Local Similarity 61.7%; Pred. No. 1.5e-51;
Matches 393; Conservative 0; Mismatches 244; Indels 0; Gaps 0;

Qy	2	TCGATACAAATAGTTTATTAATCCATGCTATCTGCTGGGTTCGATTCATCTCGTCTCT 61
Db	5166	TTATCTATTTCTTCTTCTCAATGCCAGCTATCTTTTAGGGGCAATTCATTTGGCTTAT 5107
Qy	62	GATTTGGACAACTATCTTTCAATCAATCTACGGGACATGGTCTGCTAACACTGGAA 121
Db	5106	GGATTTGGCAAGATTTTTCAAAAAATCTCCATGACTATGGCTCAGCAATCTGGAA 5047
Qy	122	GCACCAACACCTTCGCGATTTTAGTAAGAAGCTGGTATGCCAACCTTTTGTGATGACT 181
Db	5046	CAACCAATACTTTTGAATCTAGGTGTAAAGGCGAGGAATTCAGTTTTCCTTTTGACT 4987
Qy	182	TTTTCAAGGACCTAGCAAGCTGCTCCGATTTATTTTCATCTACAGGCGTTTCTC 241
Db	4986	TACTAAAGGGGACACTTCTACTCTTTTACCCCTTTTCTTCCATATCAATGGTGTTCAC 4927
Qy	242	CTCTCATCTTTGGACTTTTGGCTGTATCGGCATACCTTCCCTATCTTTTGCAGGATTTA 301
Db	4926	CTCTAATTTTCGGTCTCTTCTGCTGCTCATGTCATCTTTTCAATATTCACCGCTTTA 4867
Qy	302	AAGTGGTAAGGCTTCGCAACCACTGCTGGAGTGAATTTTCGGATTTTCGCGCTATCTTCT 361
Db	4866	AAGTGGTAAGGCTTCGCAACTCTGCTGGTGAATTTTCGGATTTTCGCGCTATCTTCT 4807
Qy	362	GTCCTACTCTGCGATATCTCTTTGGAGCTCTATCTTGGGAGTATGATTTTCACTGT 421
Db	4806	TAATTTACTCTTCTGCTCTTATCATGTCCTATGCTTTTATGAGCATGATTTAGTCTCT 4747
Qy	422	CTAGTGTACAGCATCGATTCGGCTGTATTCGGGGTCTGCTCTTTTCCACTTTTGGTT 481
Db	4746	CTAGTGTGATGCTGTATTTGCTTTATTTGGGAATCTTATTTTCCCTAGCATCGAT 4687
Qy	482	TTATCCTGAGTAACATGACTCTCTCTCATCGCTATTTATCTTATGCACTTCTAGTTTGA 541
Db	4686	TTATTTCTTACTTCTTACGACCTTTTATTTCTCAATTTATTTTGTGTTTATGCAATTTA 4627
Qy	542	TTATCATTCGTCATAGGACATAGCTCTCTATCAAAATAAATAAATAAATAAATAAATAA 601
Db	4626	TTATTTTACGACATAGACAAATCTCAAAAGAAATCAAAATAAATAAATAAATAAATAA 4567
Qy	602	CTTGGGGATTGAACCTAACCCATCAAGATCCTAAAAA 638
Db	4566	CATTTGGTTTAACTTTATCAAAACAAAGAAATA 4530

RESULT 11
LOCUS SGO236899 771 bp DNA linear BCT 05-APR-1999
DEFINITION Streptococcus gordonii parE (partial) gene and ORF2 (partial).
ACCESSION AJ236899
VERSION AJ236899.1 GI:4584071
KEYWORDS hypothetical protein; ORF2; parE gene; topoisomerase IV.
SOURCE Streptococcus gordonii.
ORGANISM Streptococcus gordonii.
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE 1 (bases 1 to 771)
AUTHORS Vriesema,A.J., Dankert,J. and Zaai,S.A.
TITLE Isolation and characterization of promoter regions from Streptococcus gordonii CH1
JOURNAL unpublished
REFERENCE 2 (bases 1 to 771)
AUTHORS Vriesema,A.J.M.
TITLE Direct Submission
JOURNAL Submitted (08-FEB-1999) Vriesema A.J.M., Medical Microbiology, Academic Medical Center, University of Amsterdam, Room L1-160,

Meibergdreef 15, 1105 A2 Amsterdam, NETHERLANDS	
FEATURES	Location/Qualifiers
source	1. .771
gene	/organism="Streptococcus gordonii" /strain="CH1" /db_xref="taxon:1302" complement(1..176) /gene="parE" complement(<1..104) /gene="parE" /codon_start=1 /transl_table=11 /product="topoisomerase IV" /protein_id="CAB40548.1" /db_xref="GI:4584072" /translation="MAKKEININYNDDAIOVLEGLDAVRKRPQWYIGS" complement(111..116) /gene="parE" complement(150..155) /gene="parE" complement(171..176) /gene="parE" 202..207 /note="ORF2" 224..229 /note="ORF2" 267..>771 /note="ORF2" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="CAB40549.1" /db_xref="GI:4584073" /translation="MINTILGLILAYLLGSIPITGLWIGIIFKKNLREYSGNMTGTTN TFRLGKTAGTDFADFLKGLTATLLPLFLHNGISPMIFGLIYVGHFPFIPAEFK GGRAVATSGVWAGFSPFLFSYLIIFIVTLXGLSMISLASIVVAGFAILSVLIFPLL GIILPSYD"
RBS	
-10_signal	
-35_signal	
-35_signal	
-10_signal	
CDS	
BASE COUNT 215 a 147 c 121 g 287 t 1 others	
ORIGIN	
Query Match 38.1%; Score 244.6; DB 1; Length 771;	
Best Local Similarity 68.0%; Pred. No. 5.7e-51;	
Matches 340; Conservative 0; Mismatches 160; Indels 0; Gaps 0;	
QY	1 ATGATTCAATAGTTTATTAATCTAGCCCTATCTGCTGGGTTCGATTCATCTGGTCTC 60
DB	270 ATAAATACAATTTTGGGCTAATTTTGGCATATTTACTGGTTCATTTCCAACTGGTTA 329
QY	61 TGGATTGGACAAGTATCTTTCAATCAATCTACGCGAGCATGCTTCTGGTAACACTGGA 120
DB	330 TGGATTGGACAAGTATCTTTTAAAAACAATTTACGACAATACGCTAGTGGAAATACGGC 389
QY	121 ACGACCAACACTTCCGCATTTTAGTAAGAAAGCTGGTATGGCAACCTTTGTGATTGAC 180
DB	390 ACAACCAACACTTTCCTGATTTTAGGAAGACACCTGGAACCTGATATCTTTGCAATCGAC 449
QY	181 TTTTTCAAAGGAACCTTAGCAACCTCTCTCCGATTTATTTTTCATCTACAAGGCGTTTCT 240
DB	450 TTTCTAAAGGAACCTTTCGCAACCTCTCTTCCACTTTTCTCCATATCAATGGTATATCA 509
QY	241 CCTCTCATCTTTGGACTTTTGGCTGTTATCGGCCATACCTTCCCTATCTTTTGGCAGGATTT 300
DB	510 CCCATGATTTTGGTCTTAATAGCTGTCTTGGCCACACTTTCCCAATTTTTCCTGGAATTT 569
QY	301 AAAGGTGTAAGGCTGTGCAACCAAGCTGCTGGAGTGAATTTTCGATTTTGGCCCTATCTTC 360
DB	570 AAGGAGGCAAGGCGGTGCTTACAAAGTGCAGGTGTGTCNTGGATTTTCTCCCTCTTC 629
QY	361 TGCTCTTACCTTGGATATCTTCTTTGGAGCTCTCTATCTTTGGCAGTATGATTTCTACTG 420
DB	630 TTTTCCCTATCTGATTTATTTATTTGTAACGCTTTTACTTGGGAAGTATGATTTCTTTG 689
QY	421 TCTAGTGTACAGCATCGATTTGGGCTGTTATCGGGGTCTGTGCTCTTTCCACTTTTTGGT 480

```

/db_xref="SPTRMBL-087723"
/translation="MTLINIKRAENDODYOMILAIWEKSVIATHDFLAAEDRFYKEQIP
LFIDLWELFWFAAGEIEVFSQTSERELDMFLFNPATGNGYGOILMIENKRINL
VDVNEQENATREYLLKHGFVSSRSLDQFGNPYPILHLQKVRNGI"
1208..1213
/gene="lmo1127"
1208..1213
/gene="lmo1127"
1220..1692
/gene="lmo1127"
1220..1675
/gene="lmo1127"
/codon_start=1
/transl_table=1
/protein_id="CAC99205.1"
/db_xref="GI:16410543"
/db_xref="SPTRMBL-087722"
/translation="MEYEOLLONELGELEIIOBELONKEYRGHIVTKAKDGTITRPL
AKKTDKAGYFVFAAREDFKNGCPFDENKADFLCIVVIGDGLNGLFFVFPCLLDIK
GILTSEASTGKMAAREPYTWCODLNKTKAKTQAMQOQVYFNKYSIRKEDH"
1660..1692
/gene="lmo1127"
complement(1689..2711)
/gene="lmo1128"
complement(1689..2711)
/gene="lmo1128"
/codon_start=1
/transl_table=1
/protein_id="CAC99206.1"
/db_xref="GI:16410544"
/db_xref="SPTRMBL-087721"
/translation="MAKNQVDTIKKWKIITIIMSALIVLVAIFIGNNYEMNETRVA
DITGGKLSAVVTPPKHHPKPGKIIIVFHCDGAOETONGYKPLMERFAKOGYISVSW
DKTGLGKSSGNLWMSDRANEVNOVLEWKKVYPDSTAKIGLWGSAGOAGWVVPKAM
NANNEIDFSLAAPAINWRQSEYNTCQAKDGTGKTKEVKQAOONELTOSALISNKK
TYESYKONGKEEMTPGRTYNFIRKNNMODATLDSLVKVKIYILALAEKKNVDSAEFK
DVYTGKIKENLETGRTYNPVSHQMINPAIANSELTNTNIVGLWVPKYFLVDNAYLDYCE
QWVSKQ"
complement(2666..3280)
/gene="lmo1129"
complement(2666..3280)
/gene="lmo1129"
/notes="similar to unknown proteins"
/codon_start=1
/transl_table=1
/protein_id="CAC99207.1"
/db_xref="GI:16410545"
/db_xref="SPTRMBL-087720"
/translation="MTLSKISDSFEKOSIVEEVASLVFAGFKSKTKTKTLFKETEADH
ITYAFCHYITERGENRYIAOIELVCGCLFVTSKSDSYQRLYQTLRKRFLFSQCLKL
IFLLSLSHXKPVONERYIDFIVTSPNFRGOGIGTKTLISCHLETFHDEKILTYVAKKNS
GAYKLERNLGFQVIOKENSTLMGVLGTGKEWKHEWOLNMGKK"
complement(3031..3996,4004..4009)
/gene="lmo1130"
complement(3277..3996)
/gene="lmo1130"
/notes="similar to transcription regulators"
/codon_start=1
/transl_table=1
/protein_id="CAC99208.1"
/db_xref="GI:16410546"
/db_xref="SPTRMBL-087719"
/translation="MTLSTGEIAOLFQISKYIRHYIDEGILVPRKPNPENGYEYEE
ADIVRLQIILFRKIGFSQIEKSELSEKVTPLMEQAEQDLOOKIEELVEIQKTOK
IYHSCWCKRNEITFVAKODRIFKVPVNOIDHNTINTSQAKHNLDPDLEPFTYFSK
IQTQGVCKLSIKRSDYTPPAGEYACKSFIAKDEATIEQTQLFLEELKYPTTSMILY
ENIYPSLAYPDANVYTYEVL"
complement(3288..3293)
/gene="lmo1129"
complement(3288..3293)
/gene="lmo1129"
complement(4004..4009)
/gene="lmo1130"

```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 24, 2002, 19:31:45 ; Search time 150.797 Seconds
(without alignments)
9587.638 Million cell updates/sec

Title: US-10-068-080-2
Perfect score: 642
Sequence: 1 atgattacaatagttttatt.....atcaagatcctaaaaataa 642

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	642	100.0	642	21	AAA09181
2	642	100.0	6171	19	AAV52170
3	640.4	99.8	642	23	AAV55594
4	640.4	99.8	642	23	AAV55801
5	327.8	51.1	344	20	AAZ20352
6	284	44.2	636	24	ABN67449
7	276.2	43.0	621	24	ABN70980
8	272.6	42.5	720	24	ABN67450
9	193.2	30.1	6691	20	AAI13025

10	124	19.3	495269	24	ABQ67195
11	102.4	16.0	615	24	ABN92939
12	102.4	16.0	818	18	AAV75059
13	102.4	16.0	3014	22	AAH54476
14	102.4	16.0	4105	22	AAH54115
15	100	15.6	582	20	AAZ20371
16	100	15.6	582	21	AAA09182
17	99.2	15.5	606	23	AAV51668
18	99.2	15.5	606	23	AAV54402
19	96	15.0	594	22	AAH53731
20	67.4	10.5	600	23	AAV53263
21	67.4	10.5	1830121	17	AAV42063
22	51	9.5	1033	20	AAV38290
23	59	9.2	570	23	AAV54065
24	57.8	9.0	612	23	AAV56022
25	56.6	8.8	189	23	AAV50887
26	53.6	8.3	618	23	AAV52569
27	49.4	7.7	1194	23	AAV56672
28	49.4	7.7	1194	23	AAV54267
29	40.4	6.3	5059	20	AAV84332
30	39	6.1	7758	24	ABL33103
31	38.6	6.0	10369	22	AAV46304
32	38.6	6.0	10369	24	ABL32393
33	38.4	6.0	618	23	AAV53979
34	38.4	6.0	16602	24	ABN80069
35	38.4	6.0	16602	24	ABL32727
36	38.2	6.0	13606	22	AAV54557
37	38.2	6.0	13606	22	AAV46561
38	38.2	6.0	13606	24	ABL33810
39	38.2	6.0	13606	24	ABK28313
40	38	5.9	781	24	ABQ14544
41	38	5.9	781	24	ABQ14545
42	37.6	5.9	522	21	AAV53681
43	37.4	5.8	32186	22	AAV34422
44	37.4	5.8	102634	21	AAV1464
45	37.4	5.8	349980	21	AAV21609

ALIGNMENTS

RESULT 1
AAA09181
ID AAA09181 standard; DNA; 642 BP.
XX AAA09181;
AC AAA09181;
XX 10-AUG-2000 (first entry)
DT
DE
DE S. pneumoniae S-ynes coding sequence.
XX S-ynes; survival; antibacterial; inhibitor; ds.
XX Streptococcus pneumoniae.
XX WO200020527-A1.
XX 13-APR-2000.
XX 30-SEP-1999; 99WO-US22665.
XX 30-SEP-1998; 98US-0163445.
XX (MILL-) MILLENNIUM PHARM INC.
XX Fritz C, Youngman P, Guzman L;
XX WPI: 2000-303799/26.
XX P-PSDB; AA192246.
XX Methods for identifying an antibacterial agent for treating
XX Streptococcus pneumoniae infections comprises detecting an interaction
XX between a ynes polypeptide and a test compound

Listeria innocua c	Staphylococcus epi	Staphylococcus aur	S. epidermidis gen	S. epidermidis gen	B. subtilis B-ynes	B. subtilis B-ynes	Staphylococcus aur	Staphylococcus aur	S. epidermidis ope	Haemophilus influe	L. lactis DNA cont	Pseudomonas aerugi	Salmonella typhi D	Staphylococcus aur	E. coli DNA for ce	DNA encoding novel	Stealth virus nucl	Human immune syste	Tumour suppressor	Human immune syste	Klebsiella pneumon	Human chemically m	Human immune syste	Chemically pretrea	Tumour suppressor	Human immune syste	DNA transcripction	Oligonucleotide fo	Oligonucleotide fo	Neisseria meningit	Human DNA for a no	N. meningitidis pa	Neisseria meningit
--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	-------------------	--------------------	--------------------	--------------------	--------------------	--------------------	-------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------	--------------------

XX Disclosure: Fig 1; 65pp; English.

XX Streptococcus pneumoniae yneS gene encodes a polypeptide (S-yneS)

CC essential for survival for a wide range of bacteria. Identifying an

CC antibacterial agent comprises contacting a yneS polypeptide (S-yneS) with

CC a test compound and detecting an interaction of the test compound with

CC the S-yneS polypeptide which indicates that the compound is an

CC antibacterial agent. Alternatively, detecting a decrease in function of

CC the polypeptide contacted with the test compound and determining whether

CC the compound inhibits growth of bacteria, relative to the growth of

CC bacteria cultured in the absence of a test compound where inhibition of

CC growth indicates the compound is an antibacterial agent. Inhibitors of

CC S-yneS function are useful for treating a Streptococcus pneumoniae

CC infection in mammals.

XX

XX Sequence 642 BP; 151 A; 143 C; 122 G; 226 T; 0 other;

Query Match 100.0%; Score 642; DB 21; Length 642;

Best Local Similarity 100.0%; Pred. No. 3.8e-178; Indels 0; Gaps 0;

Matches 642; Conservative 0; Mismatches 0;

QY 1 ATGATTACAATAGTTTATTAATCTAGCTATCTGCTGGTTCGATCCATCTGCTCTC 60

DB 1 ATGATTACAATAGTTTATTAATCTAGCTATCTGCTGGTTCGATCCATCTGCTCTC 60

QY 61 TGGATTGGACAAGTATCTTTCAATCAATCTACGGGAGCATGGTTCTGGTAACACTGGA 120

DB 61 TGGATTGGACAAGTATCTTTCAATCAATCTACGGGAGCATGGTTCTGGTAACACTGGA 120

QY 121 ACGACCAACACCTTCGCGATTTTGTAGTAAGAAAGCTGGTATGCGCAACCTTCTGATTGAC 180

DB 121 ACGACCAACACCTTCGCGATTTTGTAGTAAGAAAGCTGGTATGCGCAACCTTCTGATTGAC 180

QY 181 TTTTCAAAAGAACCTTAGCAACGCTGCTCCGATTTATTTTCATCATCAAGCGGTTTCT 240

DB 181 TTTTCAAAAGAACCTTAGCAACGCTGCTCCGATTTATTTTCATCATCAAGCGGTTTCT 240

QY 241 CCTCTCATCTTTGGACTTTTGGCTGTATCGGCCATACCTTCCCTATCTTTCAGGATTT 300

DB 241 CCTCTCATCTTTGGACTTTTGGCTGTATCGGCCATACCTTCCCTATCTTTCAGGATTT 300

QY 301 AAGGTGGTAAGGCTGTCGCAACCACTGCTGGAGTGATTTTCGGATTTGCGCTATCTTC 360

DB 301 AAGGTGGTAAGGCTGTCGCAACCACTGCTGGAGTGATTTTCGGATTTGCGCTATCTTC 360

QY 361 TGTCTCTACCTTGGCGATTAATCTTCTTGGAGCTCTCTATCTTGGCAGTATGATTTCACTG 420

DB 361 TGTCTCTACCTTGGCGATTAATCTTCTTGGAGCTCTCTATCTTGGCAGTATGATTTCACTG 420

QY 421 TCTAGTGTACAGCATCGATTGGCGCTGTATCGGGGTCTGCTCTTCCACTTTTGGT 480

DB 421 TCTAGTGTACAGCATCGATTGGCGCTGTATCGGGGTCTGCTCTTCCACTTTTGGT 480

QY 481 TTTATCTCAGTAACATATGACTCTCTCTTCATCGCTATTTATCTTAGCACTTCTAGTTTG 540

DB 481 TTTATCTCAGTAACATATGACTCTCTCTTCATCGCTATTTATCTTAGCACTTCTAGTTTG 540

QY 541 ATTATCATCTTGGCATAAGGACAAATATAGCTCGTATCAAAAATAAAGCACTGAAAATTTGGTC 600

DB 541 ATTATCATCTTGGCATAAGGACAAATATAGCTCGTATCAAAAATAAAGCACTGAAAATTTGGTC 600

QY 601 CCTTGGGATTTGAACCTTAACCCATCAAGATCCTTAAATAA 642

DB 601 CCTTGGGATTTGAACCTTAACCCATCAAGATCCTTAAATAA 642

RESULT 2

AAV52170

ID AAV52170 standard; DNA; 6171 BP.

XX

AC

AAV52170;

XX

DT 23-OCT-1998 (first entry)

DE Streptococcus pneumoniae genome fragment SEQ ID NO:37.

XX Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;

XX computer readable medium; vaccine; pharmaceutical composition; ds.

XX Streptococcus pneumoniae.

OS WO9818931-A2.

PN 07-MAY-1998.

PD 30-OCT-1997; 97WO-US19588.

PF 31-OCT-1996; 96US-0029960.

PR (HUMA-) HUMAN GENOME SCI INC.

PA Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;

PI Kunsch CA, Rosen CA;

XX WPI; 1998-272225/24.

DR Computer-readable medium with recorded Streptococcus pneumoniae

PT polynucleotide sequences - useful in diagnostic kits and assays, and

PT pharmaceutical compositions and vaccines for Streptococcus

PT pneumoniae

XX Claim 1; Page 371-374; 1409pp; English.

PS The present invention describes a computer readable medium which has

CC the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524)

CC recorded on it, or a representative fragment or a sequence at least 95%

CC identical to SEQ ID NO:1 to 391. The nucleotide sequences depicted in

CC SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from

CC Streptococcus pneumoniae. The present invention also describes an

CC isolated nucleic acid molecule encoding a homologue of any of the

CC fragments of the S. pneumoniae genome (SEQ ID NO:1 to 391) where the

CC nucleic acid molecule is produced by a process comprising: (a) screening

CC a genomic DNA library using as a probe a target sequence defined by any

CC of the sequences in SEQ ID NO:1 to 391, identifying members of the

CC library which contain sequences that hybridize to the target sequence and

CC isolating the nucleic acid molecules from the members; or (b) isolating

CC mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid

CC molecules whose nucleotide sequence is homologous to amplification

CC primers derived from the fragment of the S. pneumoniae genome to prime

CC the amplification and isolating the amplified sequences. The computer

CC readable medium can be used in a computer-based system for identifying

CC fragments of the S. pneumoniae genome of commercial importance, or

CC expression modulating fragments of the S. pneumoniae genome. Products

CC from the present invention can be used in diagnosis kits and assays, and

CC pharmaceutical compositions and vaccines for S. pneumoniae.

XX Sequence 6171 BP; 1645 A; 1424 C; 1188 G; 1914 T; 0 other;

QY Query Match 100.0%; Score 642; DB 19; Length 6171;

Best Local Similarity 100.0%; Pred. No. 9.8e-178;

Matches 642; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGATTACAATAGTTTATTAATCTAGCTATCTGCTGGTTCGATCCATCTGCTCTC 60

DB 5149 ATGATTACAATAGTTTATTAATCTAGCTATCTGCTGGTTCGATCCATCTGCTCTC 5208

QY 61 TGGATTGGACAAGTATCTTTCAATCAATCTACGGGAGCATGGTTCTGGTAACACTGGA 120

DB 5209 TGGATTGGACAAGTATCTTTCAATCAATCTACGGGAGCATGGTTCTGGTAACACTGGA 5268

QY 121 ACGACCAACACCTTCCGATTTTGTAGTAAGAAAGCTGGTATGCGCAACCTTGTGATTGAC 180

DB 5269 ACGACCAACACCTTCCGATTTTGTAGTAAGAAAGCTGGTATGCGCAACCTTGTGATTGAC 5328

QY 181 TTTTCAAAAGAACCTTAGCAACGCTGCTCCGATTTATTTTCATCTACAGCGGTTTCT 240

Db 5329 TTTTCAAGAACCCCTAGCAAGCTGCTCCGATTATTTTTCATCTACAAGCGTTTCT 5388
Qy 241 CCTCTCATCTTTGGACTTTTGGCTGTATCGGCATACCTCCCTATCTTTGCGAGGATT 300
Db 5389 CCTCTCATCTTTGGACTTTTGGCTGTATCGGCATACCTCCCTATCTTTGCGAGGATT 5448
Qy 301 AAAGTGCTAAGGCTGTCGCAACCAAGTCTCGAGTGATTTTTCGGATTTCGCCCTATCTTC 360
Db 5449 AAAGTGCTAAGGCTGTCGCAACCAAGTCTCGAGTGATTTTTCGGATTTCGCCCTATCTTC 5508
Qy 361 TGTCTCTACCTTGGCATTATCTTTTGGAGCTCTCTATCTTGGCAGTATGATTTCACTG 420
Db 5509 TGTCTCTACCTTGGCATTATCTTTTGGAGCTCTCTATCTTGGCAGTATGATTTCACTG 5568
Qy 421 TCTAGTGTACAGCATCATGATTCGGCTGTTATCGGGGTTCTGCTCTTTTCCACTTTTGGT 480
Db 5569 TCTAGTGTACAGCATCATGATTCGGCTGTTATCGGGGTTCTGCTCTTTTCCACTTTTGGT 5628
Qy 481 TTTATCTCTGAGTAACATATGACTCTCTCTTCATCGCTATTATCTTAGCACCTTGCTAGTTTG 540
Db 5629 TTTATCTCTGAGTAACATATGACTCTCTCTTCATCGCTATTATCTTAGCACCTTGCTAGTTTG 5688
Qy 541 ATTATCATCTGTCATAGGACAATATAGCTCGTATCAAAAAATAAACTGAAAAATTGGTC 600
Db 5689 ATTATCATCTGTCATAGGACAATATAGCTCGTATCAAAAAATAAACTGAAAAATTGGTC 5748
Qy 601 CCTTGGGATTGAACCTAACCCATCAAGATCCTTAAAAAATAA 642
Db 5749 CCTTGGGATTGAACCTAACCCATCAAGATCCTTAAAAAATAA 5790

RESULT 3

AAS55594

ID AAS55594 standard; DNA: 642 BP.

XX AC AAS55594;

XX DT 13-FEB-2002 (first entry)

XX DE Streptococcus pneumoniae DNA for cellular proliferation protein #165.

XX KW Antisense; ds: prokaryotic cellular proliferation gene;

XX KW antibiotic; antibacterial; drug design.

XX OS Streptococcus pneumoniae.

XX PN WO200170955-A2.

XX PD 27-SEP-2001.

XX PF 21-MAR-2001; 2001WO-US09180.

XX XX 21-MAR-2000; 2000US-191078P.

XX PR 23-MAY-2000; 2000US-206848P.

XX PR 26-MAY-2000; 2000US-207727P.

XX PR 23-OCT-2000; 2000US-242578P.

XX PR 27-NOV-2000; 2000US-253625P.

XX PR 22-DEC-2000; 2000US-257931P.

XX PR 16-FEB-2001; 2001US-269308P.

XX PA (ELIT-) ELITRA PHARM INC.

XX PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

XX PI Yamamoto RT, Xu HH;

XX DR WPI; 2001-611495/70.

XX DR P-PSDB; AAU37735.

XX PT New polynucleotides for the identification and development of

XX PT antibiotics, comprise sequences of antisense nucleic acids -

XX PS Claim 27; Seq ID No 9231; 511pp; English.

XX The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, themselves and in the discovery of novel antibiotics, the essential
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence encodes an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 642 BP; 151 A; 144 C; 122 G; 225 T; 0 other;

Query Match 99.8%; Score 640.4; DB 23; Length 642;

Best Local Similarity 99.8%; Pred. No. 1.1e-177;

Matches 641; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGATTACAATAGTTTATTAATCCTAGCCTATCTGCTGGTTCGATTCATCTGGTCTC 60

Db 1 ATGATTACAATAGTTTATTAATCCTAGCCTATCTGCTGGTTCGATTCATCTGGTCTC 60

Qy 61 TGGATTGGACAAGTATCTTTCAATCAATCTACGCGAGCATGTTCTGGTAAACATCGGA 120

Db 61 TGGATTGGACAAGTATCTTTCAATCAATCTACGCGAGCATGTTCTGGTAAACATCGGA 120

Qy 121 ACACCAACACCTTCCGCAATTTTAGTGAAGAAAGCTGTATGGCAACCTTTGTGATTGAC 180

Db 121 ACACCAACACCTTCCGCAATTTTAGTGAAGAAAGCTGTATGGCAACCTTTGTGATTGAC 180

Qy 181 TTTTCAAGAACCCCTAGCAACGCTGCTCCGATTATTTTTCATCTACAAGGCGTTTCT 240

Db 181 TTTTCAAGAACCCCTAGCAACGCTGCTCCGATTATTTTTCATCTACAAGGCGTTTCT 240

Qy 241 CCTCTCATCTTTGGACTTTTGGCTGTATCGCCATACCTTCCTATCTTTGCAAGGATT 300

Db 241 CCTCTCATCTTTGGACTTTTGGCTGTATCGCCATACCTTCCTATCTTTGCAAGGATT 300

Qy 301 AAAGTGGTAAGGCTGTCGCAACCAAGTCTCGAGTGATTTTCGGATTTCGCCCTATCTTC 360

Db 301 AAAGTGGTAAGGCTGTCGCAACCAAGTCTCGAGTGATTTTCGGATTTCGCCCTATCTTC 360

Qy 361 TGTCTCTACCTTGGCAATTATCTTTTGGAGCTCTCTATCTTTGGCAGTATGATTTCACTG 420

Db 361 TGTCTCTACCTTGGCAATTATCTTTTGGAGCTCTCTATCTTTGGCAGTATGATTTCACTG 420

Qy 421 TCTAGTGTACAGCATCGATTTCGGCTGTTATCGGGGTTCTGCTCTTTCCACTTTTGGT 480

Db 421 TCTAGTGTACAGCATCGATTTCGGCTGTTATCGGGGTTCTGCTCTTTCCACTTTTGGT 480

Qy 481 TTTATCTCTGAGTAACATATGACTCTCTCTTCATCGCTATTATCTTAGCACTTTGCTAGTTTG 540

Db 481 TTTATCTCTGAGTAACATATGACTCTCTCTTCATCGCTATTATCTTAGCACTTTGCTAGTTTG 540

Qy 541 ATTATCATCTGTCATAGGACAATATAGCTCGTATCAAAAAATAAACTGAAAAATTGGTC 600

Db 541 ATTATCATCTGTCATAGGACAATATAGCTCGTATCAAAAAATAAACTGAAAAATTGGTC 600

Qy 601 CCTTGGGATTGAACCTAACCCATCAAGATCCTTAAAAAATAA 642

Db 601 CCTTGGGATTGAACCTAACCCATCAAGATCCTTAAAAAATAA 642

RESULT 4

AAS55801
ID AAS55801 standard; DNA; 642 BP.
XX AC AAS55801;
XX XX
DT 13-FEB-2002 (first entry)
XX DE Streptococcus pneumoniae DNA for cellular proliferation protein #372.
XX KW Antisense; ds; prokaryotic cellular proliferation gene;
XX KW antibiotic; antibacterial; drug design.
XX OS Streptococcus pneumoniae.
XX PN WO200170955-A2.
XX PD 27-SEP-2001.
XX PF 21-MAR-2001; 2001WO-US09180.
XX PR 21-MAR-2000; 2000US-191078P.
XX PR 23-MAY-2000; 2000US-206848P.
XX PR 26-MAY-2000; 2000US-207727P.
XX PR 23-OCT-2000; 2000US-242578P.
XX PR 27-NOV-2000; 2000US-253625P.
XX PR 22-DEC-2000; 2000US-257931P.
XX PR 16-FEB-2001; 2001US-269308P.
XX PA (ELIT-) ELITRA PHARM INC.
XX PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX DR WPI; 2001-611495/70.
XX DR P-PSDB; AAU37942.
XX PT New polynucleotides for the identification and development of
XX PT antibiotics, comprise sequences of antisense nucleic acids -
XX PS Claim 27; Seq ID No 9438; 511pp; English.
XX CC The invention relates to antisense inhibitors of genes essential to
XX CC prokaryotic cellular proliferation, their use in identifying the
XX CC genes, their use in the discovery of novel antibiotics, the essential
XX CC genes themselves and the encoded proteins. The prokaryotes used are
XX CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
XX CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
XX CC invention is also useful for the identification of potential new targets
XX CC for antibiotic development. The antisense nucleic acids can also be used
XX CC to identify proteins used in proliferation, to express these proteins,
XX CC and to obtain antibodies capable of binding to the expressed proteins.
XX CC The proteins can be used to screen compounds in rational drug discovery
XX CC programmes. The antisense nucleic acid sequence is also useful to screen
XX CC for homologous nucleic acids which are required for cell proliferation in
XX CC a wide variety of organisms. The present sequence encodes an
XX CC essential prokaryotic cellular proliferation protein.
XX CC Note: The sequence data for this patent did not form part
XX CC of the printed specification, but was obtained in electronic
XX CC format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 642 BP; 151 A; 144 C; 122 G; 225 T; 0 other;
Query Match 99.8%; Score 640.4; DB 23; Length 642;
Best Local Similarity 99.8%; Pred. No. 1.1e-177;
Matches 641; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 ATGATTACAAGTATTTTAAATCTAGCTATCTGCTGGCTGCATTCCTCCATCTGGCTC 60
DB 1 ATGATTACAAGTATTTTAAATCTAGCTATCTGCTGGCTGCATTCCTCCATCTGGCTC 60
OY 61 TGGATTGGACAAGTATTTTCAATCAATCTAGCGAGCATGGTTCTGCTAAACACTGGA 120
|||||

Db 61 TGGATTGGACAAGTATTTTCAATCAATCTAGCGAGCATGGTTCTGCTAAACACTGGA 120
OY 121 ACGACCAACACCTTCCGCATTTTAGTAAAGAAAGCTGGTATGGCAACCTTTGTGATTGAC 180
|||||
Db 121 ACGACCAACACCTTCCGCATTTTAGTAAAGAAAGCTGGTATGGCAACCTTTGTGATTGAC 180
OY 181 TTTTCAAGAGAACCCCTAGCAACGCTCTTCCGATTTATTTTTCATCTACAAAGGCGTTTCT 240
|||||
Db 181 TTTTCAAGAGAACCCCTAGCAACGCTCTTCCGATTTATTTTTCATCTACAAAGGCGTTTCT 240
OY 241 CCTCTCATCTTTGGACTTTTGGCTGTTTATCGGCCATACCTTCCCTATCTTTTGCAGGATTT 300
|||||
Db 241 CCTCTCATCTTTGGACTTTTGGCTGTTTATCGGCCATACCTTCCCTATCTTTTGCAGGATTT 300
OY 301 AAAGTGGTAAAGGCTGCGCAACGAGTCTGGAGTGAATTTTCGGATTTGGCGCCTATCTTC 360
|||||
Db 301 AAAGTGGTAAAGGCTGCGCAACGAGTCTGGAGTGAATTTTCGGATTTGGCGCCTATCTTC 360
OY 361 TGTCTCTACCTTGGGATTTATCTTCTTGGAGCTCTCTATCTTGGCAGTATGATTTCCTG 420
|||||
Db 361 TGTCTCTACCTTGGGATTTATCTTCTTGGAGCTCTCTATCTTGGCAGTATGATTTCCTG 420
OY 421 TCTAGTGTCAAGCATCGATTGCGGCTGTTTATCGGGGTTCTGCTCTTTTCCACTTTTGGT 480
|||||
Db 421 TCTAGTGTCAAGCATCGATTGCGGCTGTTTATCGGGGTTCTGCTCTTTTCCACTTTTGGT 480
OY 481 TTTATCTCTAGTAACTATGACTCTCTCTCATGCTATTTATCTTAGCAGCTTGTAGTTTG 540
|||||
Db 481 TTTATCTCTAGTAACTATGACTCTCTCTCATGCTATTTATCTTAGCAGCTTGTAGTTTG 540
OY 541 ATTATCATCTCGTATAAGGACAATATAGCTATGCTATCAAAAATAAACTGAAATTTGGTC 600
|||||
Db 541 ATTATCATCTCGTATAAGGACAATATAGCTATGCTATCAAAAATAAACTGAAATTTGGTC 600
OY 601 CCTTGGGATTTGAACCTTAACCCATCAAGATCCTAAAAATAAA 642
|||||
Db 601 CCTTGGGATTTGAACCTTAACCCATCAAGATCCTAAAAATAAA 642

RESULT 5
AAZ20352
ID AAZ20352 standard; DNA; 344 BP.
XX AC AAZ20352;
XX XX
DT 17-NOV-1999 (first entry)
XX DE Gene encoding bacterial general essential protein gep1493.
XX KW General essential protein; pathogenic bacteria; pathogen; inhibitor;
XX KW bacterial growth; ds.
XX OS Streptococcus pneumoniae.
XX FH Key Location/Qualifiers
XX CDS 2..343
FT CD5 /*tag= a
FT /product= gep1493
XX PN WO9933871-A2.
XX PD 08-JUL-1999.
XX PF 30-DEC-1998; 98WO-US27918.
XX PR 31-DEC-1997; 97US-0070116.
XX PA (MILL-) MILLENNIUM PHARM INC.
XX PI Youngman P, Fritz C, Murphy C, Guzman L;
XX DR WPI; 1999-430230/36.
XX DR P-PSDB; AAY22561.

Db 5424 ATGAATCGTCATTTTGTACTTGTTCCTATTATTAGTTGCAATTCCTCAGGTGTT 5365
 Qy 61 TGGATTGGACAAGTATCTTTCAATCAATCTAGCCGAGCATGTTCTGTGAACACTGGA 120
 Db 5364 TGGATTGGTAACATTTCTTTTAAAGATATACAGCAATTTGGGAGTGGAAATACAGGA 5305
 Qy 121 ACGACCAACACCTTCCGCATTTTGTAGTAAAGAAAGCTGGTATGGCAACCTTTGTGATTGAC 180
 Db 5304 ACAACCAATACATTTCTGTCTTAGGAAACCTGCCGAATACGGTATTATTAAATGGAT 5245
 Qy 181 TTTTTCAGGAACCTTAGCAACCTCTTCCGATTTTTCATCTACAAAGCGTTCT 240
 Db 5244 ATCTTGAAGGAACGTTAGCCACTTCAATACCTATTTGTGTTTACAAAGCGTGAAT 5185
 Qy 241 CCTCTCATCTTTGGACCTTTTGGCTGTATCGGCCATACCTTCCCTATCTTTGAGGATTT 300
 Db 5184 CGCTCTCTTTGGGTAGCAGCTGTTTAGGGCATACCTTCCCTATTTTGGCAATTC 5125
 Qy 301 AAGGTGGTAAGCTGTGCAACAGCTGTGGAGTGTATTCGGGTTTTCGGCTATCTTTC 360
 Db 5124 AAGGTGGTAAGCTGTGCAACAGCTGTGGAGTGTATTCGGGTTTTCGGCTATCTTTC 5065
 Qy 361 TGCTCTACCTTCGGATATCTTCTTTGGAGCTCTATCTTGGCAGTATGATTTCACTG 420
 Db 5064 TTTATTATCTCTCTTATTTTGTGATTTGTCTGTATCTAACCAGTATGGTGAGTTA 5005
 Qy 421 TCTAGTGTACACATCGATTCGGCTGTATTCGGGTTTCTGCTCTTTCACATCTTTT 477
 Db 5004 ACAAGTATGATGAGCTGTACTAATACACTTCTTCTATTTATTTTGGCTTTTCACTG 4945
 Qy 478 GGTATTTATCTAGTAACTACTGCTCTCTTCTCATCGCTATTATCTTAGCACTTGTCT 537
 Db 4944 CCAGCTATTTTACCAACATTTAATCTGTTTATTAACCGTATTTGCGATTCCTTTG 4885
 Qy 538 TTGATATCATCTGTCATTAAGCAATATAGTCTGCTATCAAAATAAACTGAAATTTG 597
 Db 4884 TTTATCTTTTTCGCTCATCTGTAACATTCAGCGAATTTAAATAAGCAACAGTCTG 4825
 Qy 598 GTCCCTTGGGATTT 611
 Db 4824 CTTTCTTTGGCTT 4811

RESULT 10
 ID ABQ67195
 XX ABQ67195 standard; DNA; 495269 BP.
 AC ABQ67195;
 XX
 DT 29-AUG-2002 (first entry)
 XX
 DE Listeria innocua contig DNA sequence #8.
 KW Antibacterial; Listeria; food contamination; mutational analysis;
 KW infection; ds.
 OS Listeria innocua.
 XX
 PN WO200228891-A2.
 XX
 PD 11-APR-2002.
 XX
 PF 04-OCT-2001; 2001WO-FR03061.
 XX
 PR 04-OCT-2000; 2000FR-0012697.
 XX
 PA (INSP) INST PASTEUR.
 PA (CNRS) CNRS CENT NAT RECH SCI.
 XX
 PI Kunst F, Glaser P;
 XX
 DR WPI; 2002-332479/37.
 XX

PT New genomic sequences from *Listeria* species, useful for detection,
 PT treatment and prevention of infection, also related polypeptides,
 PT antibodies and modulators
 XX
 PS Claim 5; SEQ ID 8; 180pp; French.
 XX
 CC The present invention relates to nucleic acid sequences
 CC (AB067198-AB0671212) from *Listeria* sp. The sequences are useful as probes
 CC and primers for identification and/or detection of *Listeria* (e.g. as
 CC contaminants in foods, or mutational analysis) and for analysis of
 CC gene expression. Proteins encoded by the nucleic acid sequences can be
 CC used to screen for compounds that modulate gene expression, replication
 CC and pathogenicity of *Listeria* (potential therapeutic agents), also for
 CC treating infections by *Listeria*, and are useful as immunogens in
 CC anti-*Listeria* vaccines.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic format
 CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 495269 BP; 159529 A; 88849 C; 93135 G; 153754 T; 2 other;

Query Match 19.3%; Score 124; DB 24; Length 495269;
 Best Local Similarity 56.1%; Pred. No. 4.8e-25;
 Matches 257; Conservative 0; Mismatches 195; Indels 6; Gaps 1;
 Qy 1 ATGATTACAAATAGTATTTTATTAATCCTAGCTATCTGCTGGTTCGATTCATCTGCTC 60
 Db 441798 ATTAATTTAATTTGCTTTCTTTGTTAGCTTATGTAATCGGTCAATACCTTTGCTT 441857
 Qy 61 TGGATTGGACAAGTATCTTTCAAAATCAATCTACGCGAGCATGGTTCGTGAACACTGGA 120
 Db 441858 TGGATCGGTGTAATTTTACAAAAAAGATATTCGTGATTTTGGTAGCGGAATTTAGGA 441917
 Qy 121 ACGACCAACACCTTCCGCATTTTAGTAGAAGAGCTGTATGGCAACCTTTCTGATTGAC 180
 Db 441918 GCTACAAACTCCTTCCGTTTGTAGGGTAAAGCGGAGCATTTGTTACAGTATGAT 441977
 Qy 181 TTTTTCAGGAACCTAGCAACGCTGCTTCCGATTTATTTTCATCTACAAAGCGTTCT 240
 Db 441978 ATTTTAAAGGTACTGTGCAACATTAATCTTCTTTTCCAACTAAATTTAATCAT 442037
 Qy 241 CCTCTCATCTT-----TGGACTTTTGGCTGTATTCGCGCCATACCTTCCCTATCTTTGCA 294
 Db 442038 CATTTCTGTTTACTTACGGGAGCATTTTCGATTTATCGGACATAGTTTCCCGCTTTTCGCA 442097
 Qy 295 GGATTTAAAGTGGTAAGCTGTGCAACAGCTGCTGAGTGATTTTCGGATTTCCGCT 354
 Db 442098 GGATTCAGAGTGGGAAGCAGTTGCTACTTCTGCTGTTATCTTCTTATGACCA 442157
 Qy 355 ATCTTCTCTCTACCTTGGATTTATCTTCTTTGGAGCTCTCTATCTTGGCAGTATGAT 414
 Db 442158 CTTTGTGTTTGTAGCAGCACTAGTCTGTTTCTTATTATTAATAAATCAGCAATATG 442217
 Qy 415 TCATCTGCTAGTGTACAGCATGATTCGCTTTTCTTATTATTAATAAATCAGCAATATG 452
 Db 442218 TCGCTTAGTTCATGATGATGGAGCAGCTTGCAGCATTAAT 442255

RESULT 11
 ABN92939
 ID ABN92939 standard; DNA; 615 BP.
 XX
 AC ABN92939;
 XX
 DT 24-JUL-2002 (first entry)
 XX
 DE Staphylococcus epidermidis ORF nucleic acid sequence SEQ ID NO:2402.
 DE Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
 KW antibacterial; gene therapy; gene; ds.
 XX
 OS Staphylococcus epidermidis.
 XX

PN US6380370-B1.
PD 30-APR-2002.
PF 13-AUG-1998; 98US-0134001.
PX 14-AUG-1997; 97US-055779P.
PR 08-NOV-1997; 97US-064964P.
PS (GENO-) GENOME THERAPEUTICS CORP.
PI Doucette-Stamm LA, Bush D;
DR WP1; 2002-381255/41.
DT P-PSDB; ABP40394.
DE Novel isolated nucleic acid encoding a Staphylococcus epidermidis
PT polypeptide, useful for diagnosing and treating bacterial infections -
PS Disclosure; SEQ ID 2402; 267pp; English.
CC ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
CC frame (ORF) nucleic acid sequences which encode the amino acid sequences
CC given in ABP35124 to ABP37960. The S. epidermidis sequences have
CC antibacterial activity and can be used in gene therapy. The sequences
CC can also be used in the diagnosis and treatment of bacterial infections,
CC particularly S. epidermidis infections. The sequences can be used to
CC screen for compounds able to interfere with the S. epidermidis life
CC cycle or inhibit S. epidermidis infection.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC USPTO web site.
XX SQ Sequence 615 BP; 185 A; 81 C; 112 G; 237 T; 0 other;

Query Match 16.08; Score 102.4; DB 24; Length 615;
Best Local Similarity 50.78; Pred. No. 6.4e-20;
Matches 293; Conservative 0; Mismatches 261; Indels 24; Gaps 1;

Oy 1 ATGATTACAAATAGTTTATTATTAATCCTAGCCCTATCTGCTGGCTTCGATTCACATCGTCTC 60
Db 7 ATGATGATCATCGTCATCTTAATCTTGATTTATCTGTTGCTGATTCCTCCCAAGCGGTTA 66
Oy 61 TGGATTGGACAAATATCTTCAATCAATCTACCGAGCATGTTCTGGTTAACTGGA 120
Db 67 ATTATTGCTAAATTTTAAAAAAGATATAGACAATACGCTAGTGGAAATACCTGGA 126
Oy 121 ACGACCAACCTTCCCGCATTTTAGTGAAGAAAGCTGTATGGCAACCTTTGTGATTGAC 180
Db 127 GCAACTACAGTTTTCGGTGTCTTGGAAAGACACGCTGGATTTATAGTTACGTTTTTATAGT 186
Oy 181 TTTTTCAAAGGACCTAGCAACGCTGCTCCGATTTATTTTTCATCTACAAGCGGTT--- 237
Db 187 ATTTTCAGGGATTTATACAGCTCTTTTTCACATGTTTCCCGATGTCATCGCGATGGT 246
Oy 238 -----TCTCCTCTCATCTTTGGACATTTTGGCTGTTATCGGCCAT 276
Db 247 GTTATAAGCACCTCTTTACAAATGGTTTAAATAGTAGGATTTGTTGCAATACTCGTCAC 306
Oy 277 ACCTTCCCTATCTTTCGAGGATTTAAAGTGTTGAAGGCTGTGCGCAACGAGTCTGGAGTG 336
Db 307 GTGTATCCAAATATCTGAAATTTAATGCGGAAAGACGATGATCCAGTGCAGGAGTT 366
Oy 337 ATTTTGGGATTTGCGGCTATCTTCTCTCTACCTTGGCATTTCTTTGGAGCTCTC 396
Db 367 GTATTAGGTGTCATCTTTTACTTCTTATCTTGGCAATTTCTTTTATGTTAGTTA 426
Oy 397 TATCTTGGCAGTATGATTTCACTGCTAGTGTACAGCATTCATTCGCGCTGTTATCGGG 456
Db 427 AAAATCTTTAAATATCTTCTTTTATCAAGTATCAITGCAACATAGTTGTGTTGTTGGT 486
Oy 457 GTTCTGCTCTTCCACCTTTTGGTTTATCTCCTGAGTATCATGACTCTCTCTTCATCGGT 516

Db 487 TCAATCATCATTCATGATTATATTTTACTGCTGTAGCGGAATGTTTCAATCATATTA 546
Oy 517 ATTTCTTAGCAGCTTGTCTAGTTTGTATTATCATCATCGTCA 554
Db 547 ATAATTCGACACAATCTAATATAGTTAGTAATTTTAA 584

RESULT 12
AAV75059/c
ID AAV75059 standard; DNA; 818 BP.
XX
AC AAV75059;
XX
DT 16-MAR-1999 (first entry)
XX
DE Staphylococcus aureus contig SEQ ID #748.
XX
Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW skin infection; surgical wound infection; scalded skin syndrome;
KW toxic shock syndrome; ds.
XX
OS Staphylococcus aureus.
XX
FH Key Location/Qualifiers
FT misc_feature 121..180
FT /*tag= a
FT /*note= these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence.
XX
PN EP786519-A2.
XX
XX 30-JUL-1997.
XX
PF 07-JAN-1997; 97EP-0100117.
XX
PR 05-JAN-1996; 96US-0009861.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;
PI Rosen CA;
XX
XX WPI; 1997-374922/35.
XX
PT Polynucleotide(s) and proteins derived from Staphylococcus aureus
PT stored on computer readable medium and used in the production of
PT anti-S.aureus vaccines
XX
PS Claim 1; Page 1643-1644; 3271pp; English.
XX
XX This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC of the invention. The DNA sequences are recorded on a computer readable
CC medium, preferably selected from a floppy or hard disk, random access
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC the S.aureus DNA sequences allows putative functions to be assigned so
CC that protein-encoding or regulatory regions of commercial, therapeutic or
CC industrial importance can be obtained. Specifically, sequences which are
CC likely to encode antigens have been identified and these polypeptides can
CC be used in a vaccine composition against S.aureus infection. The
CC polypeptides can also be used in a kit for the immunodetection of
CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC skin and surgical wound infections, scalded skin syndrome, toxic shock
CC syndrome, etc. Organisms transformed with the DNA sequences can be used
CC for recombinant production of the polypeptides. The new DNA sequences
CC (and their fragments) are useful as primers or probes for isolating
CC homologues of any of the S.aureus DNA sequences contained on the
XX computer readable medium.
XX
SQ Sequence 818 BP; 294 A; 129 C; 112 G; 223 T; 60 other;

Query Match 16.0%; Score 102.4; DB 18; Length 818;
Best Local Similarity 52.6%; Pred. No. 7.2e-20;
Matches 266; Conservative 0; Mismatches 216; Indels 24; Gaps 1;

OY 1 ATGATTACAAATAGTTTATTAATCTAGCTATCTGCTGGTTCGATTCCTCTGCTC 60
DB 775 ATGATGATAATCGTCATGTTACTACTAATGTTATCTATCTGCGGCTTCCCAAGTGGATTC 716
OY 61 TCGATTGGACAAGTATCTTCAATCAATCTACCGGAGCATGGTCTGGTAACACTGGA 120
DB 715 GTAATTGGAAATTTATTTTCAAAAGATATTAGACAATTTGGTAGTGGTAATACTGGC 656
OY 121 AGGACCAACACTTCCCGCATTTTAGTAAAGAAAGCTGTATGGCAACCTTTGTGATTCAC 180
DB 655 GCTACTAATAGCTTTAGAGTATTAGGTGCTCTCGAGGATTTCTGGTAACATTTCTTAGAT 596
OY 181 TTTTCAAGGACCTAGCAACGCTGCTCCGATTTATTTTCTATCTACAAAGCGGTTTCT 240
DB 595 ATTTTCAAGGCTTCACTAATCTGTTTCTTCCCTTTATGTTTACAAAGCGGTTTCT 240
OY 241 CCTCTC-----ATCTTGGACTTTTGGCTGTTATCGGCCAT 276
DB 535 CCTATTAGTACTTTTTTACAAATGGTTTAAATCTGGCTTATTCGCTACTTGGACAC 476
OY 277 ACCTTCCCTATCTTTGGCAGGATTTAAAGGTGTAAGGCTGTGCAACAGCTGGAGTG 336
DB 475 GTTTATCTGTTTATTTAAATTTCAAGGTGGCAAGCTGTGCAACTAGTGCAGGTGC 416
OY 337 ATTTTCCGATTTGGCCATCTCTCTCTACCTTGGGATTTATCTTCTTGGAGCTCTC 396
DB 415 GTCTTGGAGTCAATCCGATACTTTTACTAATCTTCAATTTATCTTCTTATGTTATG 356
OY 397 TATCTTGGCAGTATGTTTCACTGCTAGTGTGTCACAGCATCGATGCGGCTGTTATCGGG 456
DB 355 AAGATTTTAAATATGTTCTTTAGCAAGTATCGTTGCAAGATTTGCTGTGTTGGC 296
OY 457 GTTCTGCTCTTCCACTTTTTTGGTTT 482
DB 295 TCGCTTATCTTCAAGACTATATTT 270

RESULT 13
AAH54476/c
ID AAH54476 standard; DNA; 3014 BP.
AC AAH54476;
XX
XX
XX 03-SEP-2001 (first entry)
XX
XX S. epidermidis genomic polynucleotide sequence SEQ ID NO:3840.
XX Staphylococcus epidermidis SRI strain; infection; diagnosis;
KW vaccination; endocarditis; ds.
XX
XX Staphylococcus epidermidis.
XX WO200134809-A2.
XX
XX 17-MAY-2001.
XX
XX 09-NOV-2000; 2000WO-US30782.
XX
XX 09-NOV-1999; 99US-0164258.
XX
XX (GLAX) GLAXO GROUP LTD.
XX
XX Kimmerly WJ;
XX WPI; 2001-316495/33.
XX
XX Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
PT useful for vaccinating against infections, e.g. endocarditis -

XX
PS
XX Claim 8; Page 1482-1483; 2188pp; English.
CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
CC (I) and (II) can have antibacterial activity and therefore can be used
CC in vaccination. The nucleic acids (I) may be used to produce the
CC S. epidermidis polypeptides (II) via the production of vectors
CC containing which are used to produce hosts cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to
CC AAH55098 represent oligonucleotide sequences and primers which are used
CC in the exemplification of the present invention.
CC N.B. The present invention specifically claims all the polynucleotide
CC sequences given in the sequence listing of the present specification,
CC however the sequence listing only goes up to SEQ ID NO:4454 so even
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
CC no sequences are present for SEQ ID NO:4455 to 4464.
XX
SQ Sequence 3014 BP; 1083 A; 444 C; 510 G; 977 T; 0 other;
Query Match 16.0%; Score 102.4; DB 22; Length 3014;
Best Local Similarity 50.7%; Pred. No. 1.2e-19;
Matches 233; Conservative 0; Mismatches 261; Indels 24; Gaps 1;
OY 1 ATGATTACAAATAGTTTATTAATCTAGCTATCTGCTGGTTCGATTCCTCTGCTC 60
DB 1223 ATGATGATCATCTCATCTGTTTAACTTCTGAGTTATCTGATGGTCATTCCTCCAGCGGTTA 1164
OY 61 TCGATTGGACAAGTATCTTCAAAATCAATCTACGGCAGCATGGTCTGTGAACACTGGA 120
DB 1163 ATTTTGGTAAATATTTTAAATAAAGATATAAGACAATACGGTAGTGAATACTGGA 1104
OY 121 AGCACCACACCTTCCGCACTTTTAGTGAAGAGCTGGTATGGCAACCTTTGTGATTGAC 180
DB 1103 GCACTAACAGTTTCTGCTTCTTGAAGACAGCTGGATTTATAGTTTACGTTTATAGAT 1044
OY 181 TTTTCAAGGAACCTTAGCAACGCTGCTTCCGATTTATTTTCACTACAAGCGTT --- 237
DB 1043 ATTTTCAAGGATTTATTACAGCTCTTTTCCACTATGTTTCCAGTTCATCGGATGGT 984
OY 238 -----TCTCCTCTCATCTTTGGACITTTGGCTGTTTATCGGCCAT 276
DB 983 GTTATAAGCACCTTCTTTACAAATGTTTAAATAGTAGGATTTGTCATCTCGGTCAC 924
OY 277 ACCITCCCTATCTTTGCAGGATTTAAAGTGGTAGGCTGTGCGCAACAGCTGCTGAGTG 336
DB 923 GTGATCCAAATATATCTGAAATTTAATGGCGGAAAGCAGTAGCTACCACTGCAGGAGTT 864
OY 337 ATTTTCGATTTGGCCCTATCTCTGCTCTACCTTGGATTTATCTTCTTGGAGCTCTC 396
DB 863 GTATTAGTGTCACTCTTATTTACTTCTATCTTGGCAATTTATCTTTTAGTGTATTA 804
OY 397 TATCTTGGCAGTATGATTTTCACTGCTAGTGTACAGCATCGATTTGGGCTGTATCGGG 456
DB 803 AAAATCTTTAAATATATGTTTCTTATCAAGTATCATTCAGCAATTAGTTGTGATTTGGT 744
OY 457 GTTCTGCTCTTTCCACTTTTGGTGTATCTGCTAGTAACTATGACTCTCTCTTCATCGCT 516
DB 743 TCAATCATCATTCATGATTTATTTTACTTCTGCTTACCGGAATTTGTTCAATCATATTA 684
OY 517 ATTTTCTTAGCACCTTGTAGTTTGTATTTATCATTCATTCGTCA 554
DB 683 ATAATTCGACACAATCTAATATAGTTAGATTTTAA 646
RESULT 14
AAH54115/c

AAH54115 standard; DNA; 4105 BP.
AAH54115;
03-SEP-2001 (first entry)
S. epidermidis genomic polynucleotide sequence SEQ ID NO:3479.
Staphylococcus epidermidis SRI strain; infection; diagnosis;
vaccination; endocarditis; ds.
Staphylococcus epidermidis.
W0200134809-A2.
17-MAY-2001.
09-NOV-2000; 2000WO-US30782.
09-NOV-1999; 99US-0164258.
(GLAX) GLAXO GROUP LTD.
Kimmerly WJ;
WPI; 2001-316495/33.
Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
useful for vaccinating against infections, e.g. endocarditis -
Claim 8; Page 1047-1049; 2188pp; English.
AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
(II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
(I) and (II) can have antibacterial activity and therefore can be used
in vaccination. The nucleic acids (I) may be used to produce the
S. epidermidis polypeptides (II) via the production of vectors
containing them which are used to produce hosts cells which express the
polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
used to vaccinate subjects and to raise antibodies against the bacteria.
The polypeptides may also be used to assay for other inhibitors of their
activity and therefore identify compounds that may be used for the
treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
AAH55090 represent specifically claimed S. epidermidis genomic DNA
polynucleotide sequences from the present invention. AAH55091 to
AAH55098 represent oligonucleotide sequences and primers which are used
in the exemplification of the present invention.
N.B. The present invention specifically claims all the polynucleotide
sequences given in the sequence listing of the present specification,
however the sequence listing only goes up to SEQ ID NO:4454 so even
though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
no sequences are present for SEQ ID NO:4455 to 4464.
SQ Sequence 4105 BP; 1478 A; 590 C; 744 G; 1293 T; 0 other;
Query Match 16.0%; Score 102.4; DB 22; Length 4105;
Best Local Similarity 50.7%; Pred. No. 1.4e-19;
Matches 293; Conservative 0; Mismatches 261; Indels 24; Gaps 1;
QY 1 ATGATACAAATAGTTTATTATATCTAGCTATCTGCTGGGTTCGATCCATCTCGTCTC 60
DB 1557 ATGATGATCATGTCATGTTAACTTGTGATCTATCTGATGTTGGTATCCCAAGCGGTTA 1498
QY 61 TGGATTGGACAAGTATCTTTCAAATCAATCTACGCGAGCATGTTCTGGTAACATCGTA 120
DB 1497 ATTATGGTAAATATTTTAAAAAAGATATAAGACAATACGGTAGTGGAATACTGGA 1438
QY 121 ACACCAACACCTTCGCGATTTTAGTAAGAAGCTGGTATGCGAACCTTTGTGATTGAC 180
DB 1437 GCAACTAACAGTTTTCGCTCTCTTGGAGACACGAGCTGGATTTATAGTTTACGTTTAGAT 1378
QY 181 TTTTCAAGGACCTTACACGCTGCTTCGGATTTATTTTCACTACAAGCGGTT--- 237

DB 1377 AATTTCAAGGATTTATTACAGTCTTTTTCCTCACTATGTTTCCAGTTCATCCGATGTT 1318
QY 238 -----TCTCTCTCATCTTTTGGACTTTTGGCTGTTTATCGCCAT 276
DB 1317 GTATTAAGCACCTCTTTTACAATGGTTTAATAGTAGGATTTGTTTCAATACTCGGTAC 1258
QY 277 ACCTTCCCTATCTTTGCAAGGATTTAAAGGTGGTAAAGCTGTCGCAACCAAGTCTGGAGTG 336
DB 1257 GTGATCCAATATATCTGAAATTTTAATGGCGGAAAGCAGTAGTACCAGTGCAGAGTT 1198
QY 337 AATTTCCGATTTGCGCTATCTCTCTACCTTGCATATATCTTCTTGGAGCTCTC 396
DB 1197 GTATTAGGTCAATCTTATTTTACTTCTTCTGCGCAATATATCTTTTAGTGATTATTA 1138
QY 397 TATCTTGGCAGTATGATTTTCACTGTCTAGTGTCTACAGCATGATTCGGCTGTTATCGGG 456
DB 1137 AATCTTTAAATATGTTTCTTTTATCAAGTATCATTCGACCAATTAGTTGTGATTGTT 1078
QY 457 GTTCTGCTCTTCCACTTTTGGTATTCCTGAGTAACATATGACTCTCTCTTCATCGCT 516
DB 1077 TCAATCATCTTCAATGATTATTTTACTTGTCTGTTAGCGGAATTTGTTTCAATCATATA 1018
QY 517 ATTATCTTACCACTTGTCTAGTGTGATTTGATATCATTCGTCA 554
DB 1017 ATAATTCGACACAAATCTAATATAGTAGTAATTTTAA 980

RESULT 15
AAZ20371
ID AAZ20371 standard; DNA; 582 BP.
XX AC AAZ20371;
XX DT 17-NOV-1999 (first entry)
XX DE B. subtilis B-ynes protein coding sequence.
XX KW General essential protein; pathogenic bacteria; pathogen; inhibitor;
XX KW bacterial growth; B-ynes; ds.
XX OS Bacillus subtilis.
XX FH' Key Location/Qualifiers
XX CDS 1..582 /*tag= a /product= B-ynes
XX WO9933871-A2.
XX PN 08-JUL-1999.
XX PD 30-DEC-1998; 98WO-US27918.
XX PF 31-DEC-1997; 97US-0070116.
XX PR (MILL-) MILLENNIUM PHARM INC.
XX PA Youngman P, Fritz C, Murphy C, Guzman L;
XX PI WPI; 1999-430230/36.
XX DR P-PSDB; AAY22580.
XX DR Streptococcus pneumoniae general essential protein genes and proteins,
XX PT useful for identification of antibacterial agents -
XX PS Disclosure; Fig 24; 124pp; English.
XX CC This sequence encodes the Bacillus subtilis B-ynes protein. B-ynes is
XX CC related to the Streptococcus pneumoniae general essential
XX CC protein (GEP) gene of the invention. The genes encoding the GEP
XX CC polypeptides are useful molecular tools for identifying similar genes in
XX CC pathogenic microorganisms, such as pathogenic strains of Bacillus. In
XX CC addition, the operons containing genes encoding GEP and the polypeptides

Search completed: December 24, 2002, 22:43:13
Job time : 497.797 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 24, 2002, 22:16:24 ; Search time 3529.69 Seconds
(without alignments)
2945.728 Million cell updates/sec

Title: US-10-068-080-2

Perfect score: 642

Sequence: 1 atgattacaatagttttatt.....atcaagatcctaaaaataa 642

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hct:*
9: gb_estl:*
10: gb_est2:*
11: gb_hct:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	51.6	8.0	997	17	CNS005TE
2	44.6	6.9	1154	12	BF689161
3	42.6	6.6	634	9	AL513733
4	42.4	6.6	1267	12	BG114247
5	42	6.5	633	17	BH398276
6	42	6.5	772	17	BH384505

C	7	42	6.5	1058	17	AG056521
	8	41.6	6.5	670	13	BJ153618
C	9	41.2	6.4	500	9	AU086536
C	10	41.2	6.4	1101	17	CNS00399
C	11	41	6.4	827	17	CNS02156
C	12	41	6.4	838	17	CNS03W9J
C	13	41	6.4	967	17	CNS0011A
C	14	41	6.4	1015	17	CNS00H5T
C	15	40.8	6.4	796	17	AG063366
C	16	40.8	6.4	1101	17	CNS00240
C	17	40.6	6.3	868	17	AZ547264
C	18	40.6	6.3	876	17	AZ529709
C	19	40.6	6.3	877	17	AZ547883
C	20	40.6	6.3	890	17	BH135953
C	21	40.6	6.3	899	17	BH147716
C	22	40.6	6.3	957	17	AG051372
C	23	40.4	6.3	500	9	AU088092
C	24	40.4	6.3	544	13	BJ420171
C	25	40.4	6.3	1101	17	CNS00182
C	26	40.2	6.3	588	9	AA541617
C	27	40.2	6.3	789	9	AI658642
C	28	40.2	6.3	821	17	B21581
C	29	40.2	6.3	821	17	AG031754
C	30	39.8	6.2	421	13	BI670700
C	31	39.8	6.2	430	13	BI815236
C	32	39.8	6.2	503	9	AL513809
C	33	39.8	6.2	730	12	BE866465
C	34	39.8	6.2	745	14	BQ986791
C	35	39.6	6.2	344	9	AL513975
C	36	39.6	6.2	709	17	BH402216
C	37	39.4	6.1	1089	12	BG476766
C	38	39.4	6.1	1101	17	CNS00KR2
C	39	39.2	6.1	329	9	AL513719
C	40	39.2	6.1	1058	17	CNS011HR
C	41	39	6.1	987	17	CNS00418
C	42	38.8	6.0	408	13	BM160537
C	43	38.8	6.0	431	17	CNS04J60
C	44	38.8	6.0	504	9	AL514489
C	45	38.8	6.0	510	13	BM162029

ALIGNMENTS

RESULT 1
CNS005TE
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

CNS005TE 997 bp DNA linear GSS 03-JUN-1999
Drosophila melanogaster genome survey sequence TET3 end of BAC #
BACR12K22 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
AL060767 GI:4943573
GSS.
Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 997)
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Oseogawa and
Anron Mammoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial


```

Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..633
                     /organism="Anopheles gambiae"
                     /strain="PEST"
                     /db_xref="taxon:7165"
                     /clone="AG-ND-138A17"
                     /clone_lib="ND-TAM"
                     /note="vector: pCBAC1; Site_1: HindIII"
BASE COUNT          163 a 199 c 153 g 118 t
ORIGIN
Query Match          6.5%; Score 42; DB 17; Length 633;
Best Local Similarity 64.3%; Pred. No. 0.75;
Matches 63; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

Qy 251 TTGGACTTTTGGCTGTTATCGGCATACCTTCCCTATCTTTGCAGGATTTAAAGGTGTA 310
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 561 TCGGCTCGCGCATTTATCGGCATTTGTGGCCAGTGTTTTCGGCTTTATCGCGGTA 502

Qy 311 AGGCTGTCGCAACCAAGCTGCTGGAGTGATTTTCGGATTT 348
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 501 AAGGTGTCGTACCGCTCGCGGTGTGCTGATCGGGATT 464

RESULT 7
AG056521/c
LOCUS              1058 bp DNA linear GSS 02-NOV-2001
DEFINITION        Pan troglodytes DNA, clone: PTB-042N08.F, genomic survey sequence.
ACCESSION         AG056521
VERSION           AG056521.1 GI:16593980
KEYWORDS          GSS.
SOURCE            Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
ORGANISM          Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE
  1
  Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,
  Totoki, Y., Watanabe, H. and Sakaki, Y.
  BAC end sequences of Library PTB
  JOURNAL
  Unpublished
  REFERENCE
  2 (bases 1 to 1058)
  Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,
  Totoki, Y., Watanabe, H. and Sakaki, Y.
  Direct Submission
  JOURNAL
  Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
  and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
  1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
  (E-mail:champs@psc.riken.go.jp, URL:http://hgp.psc.riken.go.jp/,
  Tel:81-45-503-9111, Fax:81-45-503-9170)
  Clones are derived from the chimpanzee BAC library PTB This BAC end
  was generated during the R&D process and may have higher chance of
  clone tracking errors.
  PRIMERS
  Sequencing: -21M13
  LIBRARY
  Vector : pKS145
  R.Site 1 : SacI
  R.Site 2 : SacI
  Location/Qualifiers
  1..1058
  /organism="Pan troglodytes"
  /db_xref="taxon:9598"
  /clone="PTB-042N08.F"
  /sex="male"
  /cell_type="lymphoblast"
  /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT          609 a 102 c 188 g 38 t 121 others
ORIGIN
Query Match          6.5%; Score 42; DB 17; Length 1058;
Best Local Similarity 45.1%; Pred. No. 0.76;
Matches 157; Conservative 0; Mismatches 188; Indels 3; Gaps 1;

Qy 206 TGCCTCGGATTTATTTTCATCTACAGCGGTTTCTCTCTCATCTTTGGACTTTTGCTG 265
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 831 TTCTTTTCTGTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 772

Qy 266 TTATCGCCATACCTTCCCTATCTTTTCAGGATTTAAAGGTGTCGACACCA 325
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 771 TTTTCTCTTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 712

Qy 326 GTGCTGGAGTGATTTTCGGATTTTCGGCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCT 385
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 711 TTCTTTTCTGTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 652

```

ACCESSION AL063914
VERSION AL063914.1 GI:4941771
KEYWORDS GSS.

063dl8 of library G from Tetraodon nigroviridis, genomic survey sequence.

ACCESSION AL263440
VERSION AL263440.1 GI:7985102
KEYWORDS GSS: genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
1 (bases 1 to 838)
Roest-crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Bouneau,L. and Weissenbach,J.
Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 838)
AUTHORS Roest-Crolius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
Bouneau,L., Billault,A., Quetier.F., Saurin,W., Bernot,A. and
Weissenbach,J.
TITLE Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis

JOURNAL Unpublished
REFERENCE 3 (bases 1 to 838)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000)
COMMENT This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
<http://www.genoscope.cns.fr/Tetraodon>.

FEATURES
Location/Qualifiers
1..838
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="063dl8"
/clone_lib="G"
/note="Genoscope sequence ID : C0BG063DB09SP1-end :
PUC-ori"

BASE COUNT 97 a 110 c 18 g 490 t 123 others

Query Match 6.4%; Score 41; DB 17; Length 838;
Best Local Similarity 35.1%; Pred. No. 1.4;
Matches 140; Conservative 47; Mismatches 212; Indels 0; Gaps 0

Qy 206 TGCCTCGCATATATTTCATACACAGGGTTTCCTCCTCATCTTGAGCATTTTGGCTG 265
Db 290 TTVTYYYYTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCTKTNTTYY 349

Qy 266 TTATCGCCCATACCTCCCTATCTTTGCAGGATTAAGGTGGTAGCTCGCAACCA 325
Db 350 TATTYTCCCTCTTTTTTYTTTTTTTTTTTTTTTTTCTTTTTTYTTTTYTTTAC 409

Qy 326 GTGCTGGAGTGATTTCGGATTTCGCCCTATCTCTGTCTACTCCGTCCGATTAFTCTCT 385
Db 410 TTCTTTYTTCTTTTCTCTCTTTTTTYTTTTYTTCTTTTTTTTTTTTTTTTTTTTT 469

Qy 386 TTGGAGCTCTCTATCTTGGCAGTATGATTTCACTGCTAGTGTACACAGCATCGATGCGG 445
Db 470 TTCCTTTTTTTTTTTTTTTTTTTTTTCTTKKKTKTTCTTCTCYKTYTGCTKKKKCTKK 529

Qy 446 CTGTTATCGGGTCTGCCTCTTCCACTTTTTTGGTTTTATCTCTGAGTAACATGACTCTC 505
Db 530 TTTK 589

Qy 506 TCTTCATCGCTATTATCTTAGTACACTTGCTAGTTTGATTATCATTCGTCATAGGACAATA 565
Db 590 TCHTTK 649

Qy 566 TAGTCGTATCAAAAATAAAACTGAAAAATTTGGTCCCCTT 604

```

DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC:
BACR34J03 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL073117.1 GI:4952850
VERSION AL073117.1
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.

REFERENCE 1 (bases 1 to 1015)
AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
TITLE Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
JOURNAL Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephyrroidea; Drosophilidae; Drosophila.
COMMENT 1 (bases 1 to 1015)
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mamoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
PI and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila\_bac.htm.
FEATURES
    source
    1..1015
        Location/Qualifiers
            /organism="Drosophila melanogaster"
            /db_xref="taxon:7227"
            /clone="BACR34J03"
            /clone_lib="RPCI-98"
            /note="end : T7"
BASE COUNT 280 a 142 c 199 g 245 t 149 others
ORIGIN
    Query Match 6.4%; Score 41; DB 17; Length 1015;
    Best Local Similarity 48.7%; Pred. No. 1.4;
    Matches 93; Conservative 6; Mismatches 92; Indels 0; Gaps 0;
QY 232 GCGGTTTCTCCCTCATCTTTGGGACTTTTGGCTGTTATCGGCCATACCTTCCCTATCTTT 291
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 203 GCGTTTTTCCCTCGTTTTTTCGGCTTTTTCCTTTTTCGCCCTTTTTCCTCGGTTT 144
QY 292 GCAGGATTTAAAGGTGGTGAAGCTGTCCGAACCACTGCTGGAGTGATTTTCGGATTGCG 351
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 143 TTCCCTTACACGGTTKTTTCTTTTTCACCTTTATTTTGTAGTCGGTGTGGTTTGGG 84
QY 352 CTAATCTCTGCTCTACCTTCGGATATCTCTTTGGAGCTCTATCTATCTGGCAGATG 411
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 83 TTTTTCGTTGCGTGTTTAAAWABACTAAATTTGGGGCGCATGTTCTCGGCATCC 24
QY 412 ATTTTCACATGTC 422
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 23 AVTTTCCCTC 13

RESULT 15
AG063366/c
LOCUS AG063366
DEFINITION Pan troglodytes DNA, clone: PTB-051P23.R, genomic survey sequence.
ACCESSION AG063366
VERSION AG063366.1 GI:16615168
KEYWORDS GSS.
SOURCE Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male
ORGANISM Pan troglodytes

```

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
1
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Tokoki, Y., Watanabe, H. and Sakaki, Y.
BAC end sequences of Library PTB
Unpublished
2 (bases 1 to 796)
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Tokoki, Y., Watanabe, H. and Sakaki, Y.
Direct Submission
Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail: chimbesegsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,
Tel: 81-45-503-9111, Fax: 81-45-503-9170)
Clones are derived from the chimpanzee BAC library PTB this BAC end
was generated during the Rad process and may have higher chance of
clone tracking errors.
PRIMERS
Sequencing: M13Rev
LIBRARY
    Vector : pKSL45
    R.Site 1 : SacI
    R.Site 2 : SacI
    Location/Qualifiers
        source
        1..796
            /organism="Pan troglodytes"
            /db_xref="taxon:9598"
            /clone="PTB-051P23.R"
            /sex="male"
            /cell_type="lymphoblast"
            /clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT 576 a 24 c 116 g 46 t 34 others
ORIGIN
    Query Match 6.4%; Score 40.8; DB 17; Length 796;
    Best Local Similarity 44.8%; Pred. No. 1.6;
    Matches 150; Conservative 0; Mismatches 185; Indels 0; Gaps 0;
QY 215 TTATTTTTCATCTACAAGCGTTTCTCTCTCATCTTTGGAGCTTTTGGCTGTATCGGCC 274
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 714 TTTCTTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 655
QY 275 ATACCTCTCCCTATCTTTCAGGATTTAAAGCTGGTGAAGCTGTCGCAACCACTGCTGGAG 334
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 654 TTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 595
QY 335 TGATTTTCGGAATTTGGCCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 394
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 594 TTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 535
QY 395 TCTATCTTGGCAGTATGATTTTCACCTGCTAGTGTACAGATCGATTCGCGCTGTATCG 454
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 534 CTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 475
QY 455 GGGTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 514
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 474 TTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 415
QY 515 CTATATCTTAGCAGCTTGTAGTTGATTAATCAAT 549
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 414 TTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 380

```

Search completed: December 25, 2002, 14:17:10
Job time : 35:36.69 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 24, 2002, 22:25:54 : Search time 93.625 Seconds
(without alignments)
2102.926 Million cell updates/sec

Title: US-10-068-080-2
Perfect score: 642
Sequence: 1 atgattacaatagttttatt.....atcaagatcctaataataa 642

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	642	100.0	6171	4	US-08-961-527-37
2	327.8	51.1	344	4	US-09-222-938A-14
3	102.4	16.0	615	4	US-09-134-001C-2402
4	100	15.6	582	4	US-09-222-938A-71
5	66	10.3	7218	1	US-08-232-463-14
6	34	5.3	521	4	US-09-221-017B-530
7	33.6	5.2	630	1	US-08-375-241-10
8	33.6	5.2	630	5	PCT-US92-06617A-10
9	31.8	5.0	2439	4	US-09-632-098-6
10	31.8	5.0	3531	2	US-08-714-402-1
11	31.6	4.9	277	4	US-09-007-005-3
12	31.6	4.9	277	4	US-09-244-796-3
13	31.6	4.9	5150	4	US-09-068-140A-14
14	31.2	4.9	99500	4	US-09-798-096-10
15	31.2	4.9	169998	4	US-09-676-610B-24
16	31	4.8	248	4	US-09-007-005-32
17	31	4.8	248	4	US-09-244-796-32
18	31	4.8	557	3	US-08-776-660A-1
19	31	4.8	624	3	US-08-776-660A-2
20	30.8	4.8	72604	4	US-09-268-992-7
21	30.8	4.8	72604	4	US-09-657-474-7
22	30.6	4.8	637	4	US-09-328-111-628
23	30.6	4.8	1074	2	US-08-975-316-54
24	30.6	4.8	1074	4	US-09-615-192A-54
25	30.6	4.8	1075	2	US-08-975-316-55
26	30.6	4.8	1075	4	US-09-615-192A-55
27	30.4	4.7	1924	4	US-08-687-590-55

Sequence 5, Appli
Sequence 17, Appl
Sequence 17, Appl
Sequence 12, Appl
Sequence 3, Appl
Sequence 24, Appl
Sequence 24, Appl
Sequence 3, Appl
Sequence 7, Appl
Sequence 1, Appl
Sequence 1, Appl
Sequence 147, App
Sequence 103, App
Sequence 360, App
Sequence 179, App
Sequence 216, App
Sequence 2, Appli

US-09-146-053-5
US-09-007-005-17
US-09-244-796-17
US-08-046-583-12
US-08-384-556A-3
US-08-331-355A-24
PCT-US94-1236A-24
PCT-US95-07753-3
US-09-157-077-7
US-09-157-077-1
US-08-678-614-1
US-09-385-222A-3
US-08-961-527-147
US-09-404-879A-103
US-08-961-527-360
US-08-961-527-179
US-08-961-527-216
US-08-926-724-2

ALIGNMENTS

RESULT 1

US-08-961-527-37
; Sequence 37, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:

CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 37:
SEQUENCE CHARACTERISTICS:
LENGTH: 6171 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-961-527-37

Query Match 100.0%; Score 642; DB 4; Length 6171;
Best Local Similarity 100.0%; Pred. No. 7.4e-193;
Matches 642; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGATTACAATAGTTTATTATTAATCTAGCCCTATCTGCTGGTTCGATTCATCTGCTC 60
|||||

Db 5149 ATGATTACAATAGTTTATTATTAATCTAGCCCTATCTGCTGGTTCGATTCATCTGCTC 5208
|||||

QY 61 TGGATTGACAAAGTATTCCTTTCAATCAATCTACCGAGGATGGTTCTGGTAAACACTGGA 120
DB 5209 TGGATTGACAAAGTATTCCTTTCAATCAATCTACCGAGGATGGTTCTGGTAAACACTGGA 5268
QY 121 ACGACCAACACTTCCCGCATTTTAGGTAAGAAAGCTGGTATGGCAACCTTTTGTGATTGAC 180
DB 5269 ACGACCAACACTTCCCGCATTTTAGGTAAGAAAGCTGGTATGGCAACCTTTTGTGATTGAC 5328
QY 181 TTTTTCAAAGAACCCCTAGCAACGCTGCTTCCGATTATTTTTCATCTACAAGGCGTTTCT 240
DB 5329 TTTTTCAAAGAACCCCTAGCAACGCTGCTTCCGATTATTTTTCATCTACAAGGCGTTTCT 5388
QY 241 CCTCTCATCTTTGGACTTTTGGCTGTTATCGGCCATACCTTCCCTATCTTTCAGGATTT 300
DB 5389 CCTCTCATCTTTGGACTTTTGGCTGTTATCGGCCATACCTTCCCTATCTTTCAGGATTT 5448
QY 301 AAAGGTGTAAGGCTGTCGCAACCAAGCTGCTGGAGTGATTTTCGGATTTCGCCCTATCTTC 360
DB 5449 AAAGGTGTAAGGCTGTCGCAACCAAGCTGCTGGAGTGATTTTCGGATTTCGCCCTATCTTC 5508
QY 361 TGTCTCTACCTTGGGATTTATCTTTTGGAGCTCTCTATCTTGGCAGTATGATTTCACTG 420
DB 5509 TGTCTCTACCTTGGGATTTATCTTTTGGAGCTCTCTATCTTGGCAGTATGATTTCACTG 5568
QY 421 TCTAGTCTCAGACATGATTCGCGCTGTTATCGGGTCTGCTTTCACACTTTTGGT 480
DB 5569 TCTAGTCTCAGACATGATTCGCGCTGTTATCGGGTCTGCTTTCACACTTTTGGT 5628
QY 481 TTTATCTCTGAGTAAGTATGACTCTCTCTTCATCGCTATTATCTTAGCACTTTCGCTAGTTG 540
DB 5629 TTTATCTCTGAGTAAGTATGACTCTCTCTTCATCGCTATTATCTTAGCACTTTCGCTAGTTG 5688
QY 541 ATTATCATCTGTCATAAGGACAATATAGCTCGTATCAAAATAAATAAATAAATAAATAA 600
DB 5689 ATTATCATCTGTCATAAGGACAATATAGCTCGTATCAAAATAAATAAATAAATAAATAA 5748
QY 601 CCTTGGGATTTGAACCTTAACCCATCAAGATCCTTAAAAATAA 642
DB 5749 CCTTGGGATTTGAACCTTAACCCATCAAGATCCTTAAAAATAA 5790

RESULT 2

US-09-222-938A-14
; Sequence 14, Application US/09222938A
; Patent No. 6437108
; GENERAL INFORMATION:
; APPLICANT: Youngman, Philip
; APPLICANT: Fritz, Christian
; APPLICANT: Murphy, Christopher
; APPLICANT: Guzman, Luz-Maria
; TITLE OF INVENTION: ESSENTIAL BACTERIAL GENES AND THEIR USE
; FILE REFERENCE: 07334/060001
; CURRENT APPLICATION NUMBER: US/09/222,938A
; CURRENT FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 14
; LENGTH: 344
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2)...(343)
US-09-222-938A-14

Query Match 51.18; Score 327.8; DB 4; Length 344;
Best Local Similarity 97.98; Pred. No. 3.3e-94;
Matches 332; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 113 ACATCGGAACCAACCACTTCCGCAATTTAGGTAAAGAAAGCTGGTATGGCAACCTTTG 172
DB 6 ACATCGGAACCAACCACTTCCGCAATTTAGGTAAAGAAAGCTGGTATGGCAACCTTTG 65

QY 173 TGATTGACTTTTTCAAAGGAACCCCTAGCAACGCTGCTTCGATTTATTTTTCATCTACAAG 232
DB 66 TGATTGACTTTTTCAAAGGAACCCCTAGCAACGCTGCTTCGATTTATTTTTCATCTACAAG 125
QY 233 GCSTTTCTCTCTCTCATCTTTGGACTTTTGGCTGTTATCGGCCATACCTTCCCTATCTTTG 292
DB 126 GCSTTTCTCTCTCTCATCTTTGGACTTTTGGCTGTTATCGGCCATACCTTCCCTATCTTTG 185
QY 293 CAGGATTTAAAGGTGGTAAGGCTGTCGCAACCAAGCTGCTGGAGTGATTTTTCGGATTTGCGC 352
DB 186 CAGGATTTAAAGGTGGTAAGGCTGTCGCAACCAAGCTGCTGGAGTGATTTTTCGGATTTGCGC 245
QY 353 CTATCTCTCTCTCTCATCTTTGGACTTTTGGCTGTTATCTTTTGGAGCTCTCTATCTTTGGCAGTATGA 412
DB 246 CTATCTCTCTCTCTCATCTTTGGACTTTTGGCTGTTATCTTTTGGAGCTCTCTATCTTTGGCAGTATGA 305
QY 413 TTTCACTGCTCTAGTGTACAGCATCGATTCGGCTGTTA 451
DB 306 TTTCACTGCTCTAGTGTACAGCATCGATTCGGCTGTTA 344

RESULT 3

US-09-134-001C-2402
; Sequence 2402, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 2402
; LENGTH: 615
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-2402

Query Match 16.08; Score 102.4; DB 4; Length 615;
Best Local Similarity 50.7%; Pred. No. 7.8e-23;
Matches 293; Conservative 0; Mismatches 261; Indels 24; Gaps 1;

QY 1 ATGATTACAACTAGTTTATTAATCCTAGCCTATCTGCTGGTTCGATTCCTCATCTGGTCTC 60
DB 7 ATGATTACAACTAGTTTATTAATCCTAGCCTATCTGCTGGTTCGATTCCTCATCTGGTCTC 66
QY 61 TGGATTGCAACAAGTATTTCTTCAAAATCAATCTACGCCAGCATGGTCTCTGTAACACTGGA 120
DB 67 ATGATTGCAACAAGTATTTCTTCAAAATCAATCTACGCCAGCATGGTCTCTGTAACACTGGA 126
QY 121 ACGACCAACACTTCCGCAATTTAGGTAAAGAAAGCTGGTATGGCAACCTTTTGTGATTGAC 180
DB 127 GCAACTAACAGTTTTCGCTGTTCTTGGAAAGACCACTGGATTTATAGTTAGCTTTTATGAT 186
QY 181 TTTTTCAAAGGAACCCCTAGCAACGCTGCTTCCGATTTATTTTTCATCTACAGGCGTTT 237
DB 187 ATTTTCAAGGATTTATTAACAGCTCTTTTTCACATGTTTTCACATGTTTTCAGTGGT 246
QY 238 -----TCTCCTCTCATCTTTGGACTTTTGGCTGTTTATCTATCTACAGGCGCAT 276
DB 247 GTTATAAGCACTTCTTTTCAAAATGGTTTAAATAGTAGGATGTTTGGCAACTACTCGGTAC 306
QY 277 ACCTTCCCTATCTTTGAGGATTTAAAGGTGGTAAAGCTCTCGCAACCACTGCTGGAGTG 336
DB 307 GTGTATCCAATATATCTGAAATTTAATGCGGAAAGCACTAGCTACCACTGCGAGGATTT 366
QY 337 ATTTTCCGATTTGCGGCTATCTCTGCTCTACCTTCCGATTTATCTTCTTTTGGAGCTCTC 396

Db	367	GTATTAGGTCGAATCCTATTATTACTTCTTATCTTGGCAATATCATCTTTTTTAGTGATTA	426
Qy	397	TATCTTTGGCAGTAGATTTTTCACCTGCTAGTGTGCACAGCATCGAATTCGGCTGTATTACGGG	456
Db	427	AAAATCTTTAAATATGTTCTTTTATCAAGTATCATTTCCAGCAATTAGTTGTGTGATTGGT	486
Qy	457	GTTCCTGCTCTTTCCACCTTTTGGTTTATACCTCGAGTAACTATGCATCTCTCTTCATCGCT	516
Db	487	TCAATCATCATTCATGATCATATATATTTTACTTGTCTTAGCGGAATGTTTCAATCATATTA	546
Qy	517	ATTATCTTAGCACTTGCTAGCTTGTGATGATATCATCTCGTCA	554
Db	547	ATAATCCACACAATCTAATATAGTAGAATTTTAA	584

RESULT 4
 US-09-222-938A-71
 : Sequence 71, Application US/09222938A
 : Patent No. 6437108
 : GENERAL INFORMATION:
 : APPLICANT: Youngman, Philip
 : APPLICANT: Fritz, Chrisian
 : APPLICANT: Murphy, Christopher
 : APPLICANT: Guzman, Luz-Maria
 : TITLE OF INVENTION: ESSENTIAL BACTERIAL GENES AND THEIR USE
 : FILE REFERENCE: 07334/060001
 : CURRENT APPLICATION NUMBER: US/09/222,938A
 : CURRENT FILING DATE: 1998-12-30
 : NUMBER OF SEQ ID NOS: 102
 : SOFTWARE: FastSeq for Windows Version 3.0
 : SEQ ID NO 71
 : LENGTH: 582
 : TYPE: DNA
 : ORGANISM: Streptococcus pneumoniae
 : FEATURE:
 : NAME/KEY: CDS
 : LOCATION: (1)...(579)
 US-09-222-938A-71

	Query Match	15.6%	Size 100;	DB 4;	Length 582;
	Best Local Similarity	52.9%	Pred. No. 4.4e-21;		
	Matches 239;	Conservative 0;	Mismatches 220;	Indels 3;	Gaps 1;
Qy	1	ATGATTACAATAGTTTATTAAATCCTACGCTATCTGCTGGGTTCGATTCCATCTGCTC	60		
Db	1	ATGTTAAATGCTTATTGATTATTTGGCTTACTTGATAGGACGATTCCTATCGGCTTA	60		
Qy	61	TGGATTGGACAAGTATTCTTTCAATCAATCTACGCGAGCATGGTTCCTGGTAAACACTGGA	120		
Db	61	ATTGGGCAAGCTTGCCAAAGGAATTCGATATTCGGGAGCAGCGAAGCGCAACTTAGGC	120		
Qy	121	ACGACCAACACTTCCCGCATTTTAGTGAAGAAGCTGCTATGGCAACCTTTGTGATTGAC	180		
Db	121	GCTACCAATGATTTCCGTACATTTGGGTAAAGAGTGGTTCGGTGCATAGCCGGAGAT	180		
Qy	181	TTTTTCAAGGAACCCCTAGCAACGCTGCTCCGATTATTTTTCATCTACAAGCGCTTCT	240		
Db	181	ATTTTGAAGGGACACTGCGCACTGCATTTGCTTTCTCATGCGATG---TGATATTGAC	237		
Qy	241	CCTCTCATCTTTGGCACTTTTGCTGTATTATCGGCCATACCTTCCCTATCTTTGCGAGATT	300		
Db	238	CCGCTTCTTGCAGGAGTCTTTGCGGTTTTTAGGCCACGTGTTTCCCATCTCGCCAAATTT	297		
Qy	301	AAAGGTGCTAAGCGTGTGCGAACCACTGCTGGAGTGAATTTCCGATTTTGCGCCTATCTC	360		
Db	298	AAAGCGGTAAGCGGTGGCGACATCAGGAGCGGTTTTGCTATTTTACGCAACCCCTGTTA	357		
Qy	361	TGCTCTACCTTGCATTATCTCTTTGGAGCTCTCTATCTATGGCAGTATGATTTCACTG	420		
Db	358	TTTATCAGATGGTTGCGGTATTCTTCATCTTTTATATCTACTGCATAAATTTGCTTCTC	417		
Qy	421	TCTAGTGTACAGCATCGATTGCGGCTGTTAT	452		

```

Db 418 TCATCGATGTTAAACAGGGATCTATACGTGTAT 449

RESULT 5
US-08-232-463-14
: Sequence 14, Application US/08232463
: Patent No. 5670367
: GENERAL INFORMATION:
: APPLICANT: DORNER, F.
: APPLICANT: SCHEIFLINGER, F.
: APPLICANT: FALKNER, F. G.
: TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
: NUMBER OF SEQUENCES: 52
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Foley & Lardner
: STREET: 1800 Diagonal Road, Suite 500
: CITY: Alexandria
: STATE: VA
: COUNTRY: USA
: ZIP: 22313-0299
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/07/935,313
: FILING DATE:
: APPLICATION NUMBER: EP 91 114 300.6
: FILING DATE: 26-AUG-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: BENT, Stephen A.
: REGISTRATION NUMBER: 29,768
: REFERENCE/DOCKET NUMBER: 30472/114 IMM
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703)836-9300
: TELEFAX: (703)683-4109
: TELEX: 899149
: INFORMATION FOR SEQ ID NO: 14:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 7218 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: IMMEDIATE SOURCE:
: CLONE: pTZgpt-Fls
US-08-232-463-14

Query Match 10.3%, Score 66; DB 1; Length 7218;
Best Local Similarity 3.1%; Pred.No. 9e-11;
Matches 12; Conservative 231; Mismatches 141; Indels 0; Gaps 0;

Qy 196 CTAGCAACGCTGCTCCGATTATTTTCATCTACAAGCGTTCTCCTCTCATGTTGCA 255
|| | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1059 CTGCGATYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1118
CTTGTGGCTGTTATCGGCATACCTCCCTATCTTTTGCAGGATTTAAAGGTGGTAAGGCT 315
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1119 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1178
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 316 GTCGCAACGAGTCGTGGAGTGNATTTCCGATTTGGCCCTATCTTCTGCTCTACCTGGC 375
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1179 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1238
ATTATCTCTTTGGAGCTCTCTATCTTGGCAGTATGATTTTCACTGCTAGTGTACAGCA 435
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1239 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1298
TCGATTCGGGCTGTTATCGGGGTTCTGCTCTTTTCCACTTTTTTGGTTTTTATCTCTGAGTAC 495

```

[illegible]

RESULT 6
US-09-221-017B-530
; Sequence 530, Application US/09221017B
; Patent No. 6444799
; GENERAL INFORMATION:
; APPLICANT: ROSS, BRUCE C.
; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
; NUMBER OF SEQUENCES: 1120
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 PAGE MILL ROAD
; CITY: Palo Alto
; STATE: CA

Query Match
5.38; Score 34; DB 4; Length 521;

[illegible]

RESULT 7
US-08-375-241-10/c
: Sequence 10, Application US/08375241
: Patent No. 5648481
: GENERAL INFORMATION:
: APPLICANT: Parodos, Kyriaki
: APPLICANT: McCarty, Janice
: TITLE OF INVENTION: Nucleic Acid Probes for the Detection of
: TITLE OF INVENTION: Shigella
: NUMBER OF SEQUENCES: 30
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
: STREET: Two Militia Drive
: CITY: Lexington
: STATE: Massachusetts
: COUNTRY: U.S.A.
: ZIP: 02173
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/375,241
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/738,800
: FILING DATE: 31-JUL-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: Granahan, Patricia
: REGISTRATION NUMBER: 32,227
: REFERENCE/DOCKET NUMBER: GTR90-04
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617-861-6240
: TELEFAX: 617-861-9540
: INFORMATION FOR SEQ ID NO: 10:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 630 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
US-08-375-241-10

	Query Match	5.28;	Score 33.6;	DB 1;	Length 630;	
	Best Local Similarity	51.7%;	Pred. No. 0.44;			
	Matches	75;	Conservative	0;	Mismatches	70; Indels 0; Gaps 0;
Oy	442	CGCGCTGTTATCCGGGGTTC	TCTCTTTCCACCTTTTTGGTTTTATCCTCAGTAACATATGAC	501		
Dd	349	GNGICGAATATCGCATTCAGCTGGTTTCTAATTTTTATCTCATCCAGTCGAAGAATTTC	290			
Oy	502	TCTCTCTTCATCGCTATTATCTTAGCACCTTGCCTAGTTTGATTATTCATTCGTCATCAAGGAC	561			

Db 289 TTTTCAACATTAGTTTAACTTCTAGTTTATTTCTATATCTTCCATCAAAATCCCTTAC 230
QY 562 AATATAGCTCTATCAAAATAAATAA 586
Db 229 AACAGGGAAGAAACATAAATAA 205

RESULT 8

PCT-US92-06617A-10/c
: Sequence 10, Application PC/TUS9206617A
: GENERAL INFORMATION:
: APPLICANT: Parodos, Kyriaki
: APPLICANT: McCarty, Janice
: TITLE OF INVENTION: Nucleic Acid Probes for the Detection of
: TITLE OF INVENTION: Shigella
: NUMBER OF SEQUENCES: 30
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Amoco Corporation
: STREET: 200 East Randolph Drive, P.O. Box 87703
: CITY: Chicago
: STATE: Illinois
: COUNTRY: U.S.A.
: ZIP: 60680

: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US92/06617A

: FILING DATE: 19920728

: PRIOR APPLICATION DATA:

: APPLICATION NUMBER: US 07/738,800

: FILING DATE: 31-JUL-1991

: ATTORNEY/AGENT INFORMATION:

: NAME: Galloway, Norval B.

: REGISTRATION NUMBER: 33,595

: REFERENCE/DOCKET NUMBER: GTR90-04 PCT

: TELECOMMUNICATION INFORMATION:

: TELEPHONE: 312-856-7180

: TELEFAX: 312-856-4972

: INFORMATION FOR SEQ ID NO: 10:

: SEQUENCE CHARACTERISTICS:

: LENGTH: 630 base pairs

: TYPE: NUCLEIC ACID

: STRANDEDNESS: double

: TOPOLOGY: linear

PCT-US92-06617A-10

Query Match 5.28; Score 33.6; DB 5; Length 630;

Best Local Similarity 51.74; Pred. No. 0.44; Mismatches 0; Indels 0; Gaps 0;

Matches 75; Conservative 0;

QY 442 GCGGCTGTATCGGGTCTCTCTTTCCACTTTTGGTATCTTATCTAGTAAGTATGAC 501
Db 349 GNGTCGATATCTGCACTGCTGTTCTAATTTTATCTATCCAGTGAAGATTC 290
QY 502 TCTCTCTTCAGCGTATATCTCTAGCACTTGTAGTTGATATCATTCGTCAAGAAC 561
Db 289 TTTTCAACATTAGTTTAACTTCTAGTTTATTTCTATATCTTCCATCAAAATCCCTTAC 230

QY 562 AATATAGCTCTATCAAAATAAATAA 586

Db 229 AACAGGGAAGAAACATAAATAA 205

RESULT 9

US-09-632-098-6/c
: Sequence 6, Application US/09632098
: Patent No. 6420154
: GENERAL INFORMATION:
: APPLICANT: Sheppard, Paul O.

: APPLICANT: Baidur, Nand
: APPLICANT: Bishop, Paul D.
: TITLE OF INVENTION: MAMMALIAN ADHESION PROTEASE PEPTIDES
: FILE REFERENCE: 99-39
: CURRENT APPLICATION NUMBER: US/09/632,098
: CURRENT FILING DATE: 2000-08-02
: NUMBER OF SEQ ID NOS: 26
: SOFTWARE: FastSEQ for Windows Version 3.0
: SEQ ID NO 6
: LENGTH: 2439
: TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: Degenerate sequence
: NAME/KEY: misc_feature
: LOCATION: (1)...(2439)
: OTHER INFORMATION: n = A,T,C or G
US-09-632-098-6

Query Match 5.0%; Score 31.8; DB 4; Length 2439;

Best Local Similarity 27.4%; Pred. No. 3.3; Mismatches 225; Indels 1; Gaps 1;

Matches 104; Conservative 49;

QY 54 TGGTCTCTGGATTGGACAAGTATTCTTTCAAAATCAATCTACGCGAGCATGTTCTGTAA 113

Db 2096 TGTFTTCNGCYTGNACNGNCCNSWRTCCATNSMNCNCCRAANCCNGGYTTRTCAR 2037

QY 114 CACTGGAACGACCAACACCTTCCGCAATTTAGGTAAAGAAAGCTGGTATGGCAACCTTTGT 173

Db 2036 AANGGNGGCCANCCNGCGRCARTTGTGRTTNSWRTTRCANACNCCRTGN 1977

QY 174 GATTGACTTTTCAAGGAACCTTAGCAACGCTCTCCGATTATTTTTCATCTACAAGG 233

Db 1976 SWRTGRCANGNGTNRANCANCKYTGNARYTCYTGRAANGCRT-TYTNKRCANCKNCK 1918

QY 234 CGTTTCTCTCTCATCTTTTGGACTTTTGGCTGTTATCGGCATACCTTCCCTATCTTGC 293

Db 1917 NSWYTGRCANACCATNCKNGNCCRCAYTGNCTNCCNGGYTCNACNARNCCNARNCCNAR 1858

QY 294 AGATTAAAGGTGTAAGGCTGTGCGAACAGCTGTGCGAGTGATTTTGGGATTTGGGCC 353

Db 1857 NARRTCNARYTGNCSWNGNARNCGNCCNCKRCANGTNACYTCYTGNCCTC 1798

QY 354 TATCTTCTCTCTACCTTGGATTATCTTCTTTGGAGCTCTCTATCTTGGCAGTATGAT 413

Db 1797 NARRTGNACNGTNSWRTCNACNGNACCATRTGNGNGCNARNARNNSWNGGYTTNCCNCC 1738

QY 414 TTCACCTGTCTAGTGTCACA 432

Db 1737 YTGRCAYTGNARYTTNCCR 1719

RESULT 10

US-08-714-402-1/c

: Sequence 1, Application US/08714402

: Patent No. 5910441

: GENERAL INFORMATION:

: APPLICANT: ROCHA, Claudia

: APPLICANT: FISCHETTI, Vincent A.

: TITLE OF INVENTION: FIBRONECTIN AND FIBRINOGEN BINDING

: TITLE OF INVENTION: PROTEIN FROM GROUP A STREPTOCOCCI

: NUMBER OF SEQUENCES: 2

: CORRESPONDENCE ADDRESS:

: ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS

: STREET: P.O. Box 1404

: CITY: Alexandria

: STATE: Virginia

: COUNTRY: United States

: ZIP: 22313-1404

: COMPUTER READABLE FORM:

: MEDIUM TYPE: Floppy disk

: COMPUTER: IBM PC compatible

: OPERATING SYSTEM: PC-DOS/MS-DOS

COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/068,140A
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP96/04807
FILING DATE: No. 6281409ember 4, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Dinner, Dara L.
REGISTRATION NUMBER: 33,680
REFERENCE/DOCKET NUMBER: C70237
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5017
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 5150 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Ribes nigrum
STRAIN: Ben Alder
US-09-068-140A-14

Query Match 4.9%; Score 31.6; DB 4; Length 5150;
Best Local Similarity 50.7%; Pred. No. 5.6; Mismatches 0; Gaps 0;
Matches 76; Conservative 0;
Qy 336 GATTTCGGATTGGCGCTATCTCTGCTCTACCTTCGCGATTATCTCTTTGGAGCTCT 395
Db 167 GATTTCGCAAAATGCTGCTACTTTAGTTAGATCATTAAGTTTCTCTAGTTTCT 108
Qy 396 CTATCTGGCAGTATGATTCACCTGCTAGTGTGCACAGCATCGATTGGCGCTGTATCGG 455
Db 107 CACAGTTCTAGTCTCTATTTCCAGTTTATCAGTTGATCGTCCATAGCATGATTTCCA 48
Qy 456 GTTCTGCTCTTCCACTTTTGGTTTAT 485
Db 47 GTTTTGATGATTTCCACCGATTGATTTAT 18

RESULT 14
US-09-798-096-10
Sequence 10, Application US/09798096
Patent No. 6399378
GENERAL INFORMATION:
APPLICANT: Donna T. Ward
APPLICANT: Andrew T. Watt
TITLE OF INVENTION: ANTISENSE MODULATION OF RECQL2 EXPRESSION
FILE REFERENCE: RTS-0207
CURRENT APPLICATION NUMBER: US/09/798,096
CURRENT FILING DATE: 2001-03-01
NUMBER OF SEQ ID NOS: 89
SEQ ID NO 10
LENGTH: 99500
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
US-09-798-096-10

Query Match 4.9%; Score 31.2; DB 4; Length 99500;

Best Local Similarity 46.4%; Pred. No. 35;
Matches 102; Conservative 0; Mismatches 118; Indels 0; Gaps 0;
Qy 289 TTTCGAGGATTTAAAGGTGGTAAGGCTGTCGCAACCAAGTCTCGAGTGTATTTTCGATTT 348
Db 57656 TGTTCAAGCTGTTTCAGTGTGTTTGGTGTGTTTGAATAAATGCTGCTGTGAACATTTGTGA 57715
Qy 349 GCSCCTATCTCTGCTCTACCTTGCGATTATCTCTTTGGAGCTCTCTATCTTTGGCAGT 408
Db 57716 GAAGTGTGCTGTCGTCATCAGTTTATTTCTCTTGGTGTGCTGGTAACTGT 57775
Qy 409 ATGATTTTCACTGTCTAGTGTCAAGCATCGATTGCGGCTGTATTCGGGGTTCGTCTCTTT 468
Db 57776 ATGATTAACCTTTTCAATTCAAGGACTACTAACTAGTTTCAAAGTGCCTCGCTCTTTT 57835
Qy 469 CCACCTTTTGGTGTATCTCTGAGTAACTATGACTCTCTCT 508
Db 57836 TTTTGTGTTTGTGTTTGTGAGCGGAGTCTCACTCT 57875

RESULT 15
US-09-676-610B-24/C
Sequence 24, Application US/09676610B
Patent No. 644465
GENERAL INFORMATION:
APPLICANT: C. Frank Bennett
APPLICANT: Jacqueline Wyatt
APPLICANT: Susan M. Freier
TITLE OF INVENTION: OLIGONUCLEOTIDE INHIBITION OF HER-1 EXPRESSION
FILE REFERENCE: RTS-0138
CURRENT APPLICATION NUMBER: US/09/676,610B
CURRENT FILING DATE: 2000-09-29
NUMBER OF SEQ ID NOS: 182
SEQ ID NO 24
LENGTH: 169998
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: exon
LOCATION: (1208)...(1472)
NAME/KEY: intron
LOCATION: (1473)...(124390)
NAME/KEY: exon
LOCATION: (124391)...(124544)
NAME/KEY: intron
LOCATION: (124545)...(125409)
NAME/KEY: exon
LOCATION: (125410)...(125595)
NAME/KEY: intron
LOCATION: (125596)...(128711)
NAME/KEY: exon
LOCATION: (128712)...(128848)
NAME/KEY: intron
LOCATION: (128849)...(133400)
NAME/KEY: exon
LOCATION: (133401)...(133469)
NAME/KEY: intron
LOCATION: (133470)...(134652)
NAME/KEY: exon
LOCATION: (134653)...(134773)
NAME/KEY: intron
LOCATION: (134774)...(136116)
NAME/KEY: exon
LOCATION: (136117)...(136261)
NAME/KEY: intron
LOCATION: (136262)...(137936)
NAME/KEY: exon
LOCATION: (137937)...(138053)
NAME/KEY: intron
LOCATION: (138054)...(138637)
NAME/KEY: exon
LOCATION: (138638)...(138766)
NAME/KEY: intron

LOCATION: (138767)...(138864)
NAME/KEY: exon
LOCATION: (138865)...(138940)
NAME/KEY: intron
LOCATION: (138941)...(139765)
NAME/KEY: exon
LOCATION: (139766)...(139860)
NAME/KEY: intron
LOCATION: (139861)...(142245)
NAME/KEY: exon
LOCATION: (142246)...(142445)
NAME/KEY: intron
LOCATION: (142446)...(143605)
NAME/KEY: exon
LOCATION: (143606)...(143738)
NAME/KEY: intron
LOCATION: (143739)...(145838)
NAME/KEY: exon
LOCATION: (145839)...(145931)
NAME/KEY: intron
LOCATION: (145932)...(147385)
NAME/KEY: exon
LOCATION: (147386)...(147544)
NAME/KEY: intron
LOCATION: (147545)...(153274)
NAME/KEY: exon
LOCATION: (153275)...(153321)
NAME/KEY: intron
LOCATION: (153322)...(155088)
NAME/KEY: exon
LOCATION: (155089)...(155231)
NAME/KEY: intron
LOCATION: (155232)...(156025)
NAME/KEY: exon
LOCATION: (156026)...(156151)
NAME/KEY: intron
LOCATION: (156152)...(156826)
NAME/KEY: exon
LOCATION: (156827)...(156928)
NAME/KEY: intron
LOCATION: (156929)...(163399)
NAME/KEY: exon
LOCATION: (163400)...(163586)
US-09-676-610B-24

Query Match 4.9%; Score 31.2; DB 4; Length 16998;
Best Local Similarity 46.1%; Pred. No. 46;
Matches 105; Conservative 0; Mismatches 123; Indels 0; Gaps 0;
QY 413 TTTCACGTCTAGTGTACACAGCATCGATTGGGGCTGTTATCGGGGTCTGCTCTTTCCAC 472
Db 167319 TCATCGTCTACTGGTCTCATATGGATCTCCCAAGACACCTCTGTTTTTTCCTTT 167260
QY 473 TTTTGGTTTTTACCTCAGTAATATGACCTCTCTCTTCATCGCTATTATCTTAGCACTTG 532
Db 167259 TTCTTTTGTCTAGTTGATATATTTTATGTAATCTCTTTGGCAGTTTAGTCAGGACAAT 167200
QY 533 CTAGTTTGTATATCATTCGTCATAAGACAATATAGCTGATCAAAAATAAACTGAAA 592
Db 167199 GAAACTTGCATAACAGTGGCCAGCTCTCCCATCAACATTTAGAAGAAAACCTTACCTAGAT 167140
QY 593 ATTGTGTCCTTGGGATTGCAACTAACCCATCAAGATCCTAAAAAT 640
Db 167139 AGTTTGTGCTTTGTGAGTCAACAGGACATTTATAAGCCATAAAT 167092

Search completed: December 25, 2002, 14:24:38
Job time : 177.625 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 25, 2002, 11:18:11 ; Search time 189.61 Seconds
(without alignments)
1375.466 Million cell updates/sec

Title: US-10-068-080-2
Perfect score: 642
Sequence: 1 atgattacaatagttttatt.....atcaagatcctaaaaataa 642

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 363474 seqs, 203117208 residues

Total number of hits satisfying chosen parameters: 726948

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications, NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	642	100.0	642	12	US-10-068-080-2
2	640.4	99.8	642	10	US-09-815-242-9231
3	640.4	99.8	642	10	US-09-815-242-9438
c 4	193.2	30.1	6691	10	US-09-070-927A-88
5	102.4	16.0	609	10	US-09-823-246-1
6	100	15.6	582	12	US-10-068-080-4
7	99.2	15.5	606	10	US-09-815-242-4250
8	99.2	15.5	609	10	US-09-815-242-8039
9	67.4	10.5	600	10	US-09-815-242-6900
10	59	9.2	570	10	US-09-815-242-7702
11	57.8	9.0	612	10	US-09-815-242-9659
c 12	56.6	8.8	189	10	US-09-815-242-3464
13	53.6	8.3	618	10	US-09-815-242-6206
14	38.4	6.0	618	10	US-09-815-242-7616
c 15	36.6	5.7	1118	9	US-09-938-842A-4649
16	36.4	5.7	358	10	US-09-783-590-5896
17	34.4	5.4	400	10	US-09-960-352-2619
18	34.4	5.4	403	10	US-09-960-352-4074
19	34.4	5.4	424	10	US-09-960-352-11631

20	34.4	5.4	430	10	US-09-960-352-9747
c 21	34	5.3	659158	9	US-09-771-208-20
22	33.2	5.2	326014	10	US-09-731-231A-3
23	32.6	5.1	408	10	US-09-960-352-1221
c 24	32.4	5.0	5621	10	US-09-070-927A-571
c 25	32.4	5.0	6855	10	US-09-764-864-1694
26	32.4	5.0	21045	10	US-09-764-864-1695
27	32.2	5.0	204	9	US-10-040-739-505
c 28	32.2	5.0	3087	10	US-09-815-242-4830
c 29	32.2	5.0	3144	10	US-09-815-242-8985
c 30	31.8	5.0	9439	10	US-09-764-877-2224
c 31	31.6	4.9	718	9	US-10-001-857-107
c 32	31.6	4.9	988	9	US-10-016-157A-29
c 33	31.6	4.9	1671	10	US-09-815-242-5959
c 34	31.6	4.9	5150	10	US-09-800-528-14
35	31.4	4.9	471	9	US-09-946-807-1424
36	31.4	4.9	471	10	US-09-795-668-1424
37	31.4	4.9	471	10	US-09-795-668-1424
38	31.4	4.9	598	10	US-09-864-761-12491
39	31.4	4.9	17493	10	US-09-804-471A-3
40	31.2	4.9	166	10	US-09-770-696-376
41	31.2	4.9	264	10	US-09-969-373-472
c 42	31.2	4.9	597	9	US-09-938-842A-766
43	31.2	4.9	717	10	US-09-770-149-172
c 44	31.2	4.9	1416	10	US-09-925-299-173
c 45	31.2	4.9	197496	9	US-09-877-177-10

ALIGNMENTS

RESULT 1
US-10-068-080-2
; Sequence 2, Application US/10068080
; Patent No. US20020115591A1
; GENERAL INFORMATION:
; APPLICANT: Fritze, Christian
; APPLICANT: Youngman, Philip
; TITLE OF INVENTION: USE OF YNES, ESSENTIAL BACTERIAL GENES AND POLYPEPTIDES
; FILE REFERENCE: 07334/090001
; CURRENT APPLICATION NUMBER: US/10/068,080
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/070,116
; PRIOR FILING DATE: 1997-12-31
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 642
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(639)
US-10-068-080-2

Query Match	100.0%	Score	642	DB	12	Length	642
Best Local Similarity	100.0%	Pred. No.	3.3e-180				
Matches	642	Conservative	0	Mismatches	0	Indels	0
0		Gaps	0				
Qy	1	ATGATTACAATAGTTTATTAACTAGCTATCTGCTGGTTCGATTCCATCTGGTCTC	60				
Db	1	ATGATTACAATAGTTTATTAACTAGCTATCTGCTGGTTCGATTCCATCTGGTCTC	60				
Qy	61	TGGATTGGACAAGTATTTCTTCAATCAATCTACCGAGCATGTTCTTGGTAACACTGGA	120				
Db	61	TGGATTGGACAAGTATTTCTTCAATCAATCTACCGAGCATGTTCTTGGTAACACTGGA	120				
Qy	121	ACGACCAACACCTTCCCGATTTAGTATGGAAGACTGGTATGGCAACCTTTGTGATTGAC	180				
Db	121	ACGACCAACACCTTCCCGATTTAGTATGGAAGACTGGTATGGCAACCTTTGTGATTGAC	180				
Qy	181	TTTTTCAAGGACCCCTAGCAACGCTCTCCGATTATTTTTCATCTACAGGGGTTTCT	240				
Db	181	TTTTTCAAGGACCCCTAGCAACGCTCTCCGATTATTTTTCATCTACAGGGGTTTCT	240				

Db 181 TTTTCAAGAACCCCTAGCAACGCTGCTCCGATTATTTTTCATCTACAGGCGTTTCT 240
Qy 241 CCTCTCATCTTTGGACATTTTGGCTGTATCGGCATACCTTCCCTATCTTTTGCAGGATTT 300
Db 241 CCTCTCATCTTTGGACATTTTGGCTGTATCGGCATACCTTCCCTATCTTTTGCAGGATTT 300
Qy 301 AAAGGTGGTAGGCTGTCGCAACGAGTGGAGTGTATTTTGGGATTTGGCGCTATCTTC 360
Db 301 AAAGGTGGTAGGCTGTCGCAACGAGTGGAGTGTATTTTGGGATTTGGCGCTATCTTC 360
Qy 361 TGTCTCATCTTTGGGATTTATCTTTTGGAGCTCTCTATCTTGGCAGTATGATTTTCACTG 420
Db 361 TGTCTCATCTTTGGGATTTATCTTTTGGAGCTCTCTATCTTGGCAGTATGATTTTCACTG 420
Qy 421 TCTAGTGTACAGCATCGATTTGGGCTGTATCGGGGTTCGCTTTTCCACTTTTGGT 480
Db 421 TCTAGTGTACAGCATCGATTTGGGCTGTATCGGGGTTCGCTTTTCCACTTTTGGT 480
Qy 481 TTTATCTCTGAGTAACTATGACTCTCTTTCATCGCTATTTAGCACTTGCCTAGTTTG 540
Db 481 TTTATCTCTGAGTAACTATGACTCTCTTTCATCGCTATTTAGCACTTGCCTAGTTTG 540
Qy 541 ATTATCATCTCTATAGGACAATATAGCTCGTATCAAAAATAAAACTGAAAAATTTGGTC 600
Db 541 ATTATCATCTCTATAGGACAATATAGCTCGTATCAAAAATAAAACTGAAAAATTTGGTC 600
Qy 601 CCTTGGGATTTGAACCTTAACCCATCAAGATCCTAAAAATAA 642
Db 601 CCTTGGGATTTGAACCTTAACCCATCAAGATCCTAAAAATAA 642

RESULT 2

US-09-815-242-9231
; Sequence 9231, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9231
; LENGTH: 642
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(642)
US-09-815-242-9231

Query Match 99.8%; Score 640.4; DB 10; Length 642;
Best Local Similarity 99.8%; Pred. No. 9.8e-180;
Matches 641; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 ATGATTACAATAGTATTTTATTAAATCCCTAGCCTATCTGCTGGTTCGATTCCCATCTGGTCTC 60
Db 1 ATGATTACAATAGTATTTTATTAAATCCCTAGCCTATCTGCTGGTTCGATTCCCATCTGGTCTC 60
Qy 61 TGAATTGGACAAGTATTTCTTCAAAATCAATCTACGCGAGCATGGTCTTGGTAACACTGGA 120
Db 61 TGAATTGGACAAGTATTTCTTTCAAAATCAATCTACGCGAGCATGGTCTTGGTAACACTGGA 120
Qy 121 ACAGCAACACCTTCCCATTTTAGGTAAAGAAAGCTGGTATGGCAACCTTTTGTGATTTGAC 180
Db 121 ACAGCAACACCTTCCCATTTTAGGTAAAGAAAGCTGGTATGGCAACCTTTTGTGATTTGAC 180
Qy 181 TTTTCAAGAAACCCCTAGCAACGCTGCTTCCGATTTATTTTTCATCTACAGGCGTTTCT 240
Db 181 TTTTCAAGAAACCCCTAGCAACGCTGCTTCCGATTTATTTTTCATCTACAGGCGTTTCT 240
Qy 241 CCTCTCATCTTTGGACTTTTGGCTGTATCGGCCATACCTTCCCTATCTTTTGCAGGATTT 300
Db 241 CCTCTCATCTTTGGACTTTTGGCTGTATCGGCCATACCTTCCCTATCTTTTGCAGGATTT 300
Qy 301 AAAGGTGGTAGGCTGTCGCAACGAGTGGAGTGTATTTTGGGATTTGGCGCTATCTTC 360
Db 301 AAAGGTGGTAGGCTGTCGCAACGAGTGGAGTGTATTTTGGGATTTGGCGCTATCTTC 360
Qy 361 TGTCTCTACCTTGGGATTTATCTTTTGGAGCTCTCTATCTTGGAGCTCTCTATCTTGGAGTGTATTTTCACTG 420
Db 361 TGTCTCTACCTTGGGATTTATCTTTTGGAGCTCTCTATCTTGGAGTGTATTTTCACTG 420
Qy 421 TCTAGTGTACAGCATCGATTTGGGCTGTATCGGGGTTCGCTTTTCCACTTTTGGT 480
Db 421 TCTAGTGTACAGCATCGATTTGGGCTGTATCGGGGTTCGCTTTTCCACTTTTGGT 480
Qy 481 TTTATCTCTGAGTAACTATGACTCTCTTTCATCGCTATTTAGCACTTGCCTAGTTTG 540
Db 481 TTTATCTCTGAGTAACTATGACTCTCTTTCATCGCTATTTAGCACTTGCCTAGTTTG 540
Qy 541 ATTATCATCTCTATAGGACAATATAGCTCGTATCAAAAATAAAACTGAAAAATTTGGTC 600
Db 541 ATTATCATCTCTATAGGACAATATAGCTCGTATCAAAAATAAAACTGAAAAATTTGGTC 600
Qy 601 CCTTGGGATTTGAACCTTAACCCATCAAGATCCTAAAAATAA 642
Db 601 CCTTGGGATTTGAACCTTAACCCATCAAGATCCTAAAAATAA 642

RESULT 3

US-09-815-242-9438
; Sequence 9438, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727

;; PRIOR FILING DATE: 2000-05-26
;; PRIOR APPLICATION NUMBER: 60/242,578
;; PRIOR FILING DATE: 2000-10-23
;; PRIOR APPLICATION NUMBER: 60/253,625
;; PRIOR FILING DATE: 2000-11-27
;; PRIOR APPLICATION NUMBER: 60/257,931
;; PRIOR FILING DATE: 2000-12-22
;; PRIOR APPLICATION NUMBER: 60/269,308
;; PRIOR FILING DATE: 2001-02-16
;; NUMBER OF SEQ ID NOS: 14110
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 9438
;; LENGTH: 642
;; TYPE: DNA
;; ORGANISM: Streptococcus pneumoniae
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (1)...(642)
US-09-815-242-9438

Query Match 99.8%; Score 640.4; DB 10; Length 642;
Best Local Similarity 99.8%; Pred. No. 9.8e-180;
Matches 641; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGATTACAAGTATTTTAAATCTAGCCTAGCTGCTGGTTCGATTCCTGCTGCTC 60
Db 1 ATGATTACAAGTATTTTAAATCTAGCCTAGCTGCTGGTTCGATTCCTGCTGCTC 60

Qy 61 TGGATTGGACAAGTATTTTAAATCAATCTACCGAGCATGGTTCGTGTAACACTGGA 120
Db 61 TGGATTGGACAAGTATTTTAAATCAATCTACCGAGCATGGTTCGTGTAACACTGGA 120

Qy 121 ACGACCAACACTTCCGCAATTTAGTAAAGAAAGCTGATGGCAACCTTTGTGATTCAC 180
Db 121 ACGACCAACACTTCCGCAATTTAGTAAAGAAAGCTGATGGCAACCTTTGTGATTCAC 180

Qy 181 TTTTCAAGGAACCTTAGCAGCTGCTCCGATATTTTCACTACAAAGCGCTTCT 240
Db 181 TTTTCAAGGAACCTTAGCAGCTGCTCCGATATTTTCACTACAAAGCGCTTCT 240

Qy 241 CCTCTCATCTTTGGACTTTTGGCTTTATCGGCCATACCTTCCCTATCTTTCAGGATTT 300
Db 241 CCTCTCATCTTTGGACTTTTGGCTTTATCGGCCATACCTTCCCTATCTTTCAGGATTT 300

Qy 301 AAAGTGTAGAGCTGTCGCAACCAAGCTGCTGGAGTATTTCCGATTTGCGCTATCTTC 360
Db 301 AAAGTGTAGAGCTGTCGCAACCAAGCTGCTGGAGTATTTCCGATTTGCGCTATCTTC 360

Qy 361 TGTCTCTACCTTGGCAATATCTTCTTGGAGCTCTCTATCTTGGCAGTATGATTTCACTG 420
Db 361 TGTCTCTACCTTGGCAATATCTTCTTGGAGCTCTCTATCTTGGCAGTATGATTTCACTG 420

Qy 421 TCTAGTGTACAGCATGATTCGCGCTGTTATCGGGGTTCTGCTCTTCCACTTTTGGT 480
Db 421 TCTAGTGTACAGCATGATTCGCGCTGTTATCGGGGTTCTGCTCTTCCACTTTTGGT 480

Qy 481 TTTATCTTGAAGTATGATTCCTCTCTCATCGCTATTTCTTACGACTTTGCTAGTTTG 540
Db 481 TTTATCTTGAAGTATGATTCCTCTCTCATCGCTATTTCTTACGACTTTGCTAGTTTG 540

Qy 541 ATTATCATCTGTCATAGGACAATATAGCTGATCAAAATAAATAAATTTTGGTCTC 600
Db 541 ATTATCATCTGTCATAGGACAATATAGCTGATCAAAATAAATAAATTTTGGTCTC 600

Qy 601 CCTTGGGATTTGAACCTTAACCCATCAAGATCCTTAAAAATAA 642
Db 601 CCTTGGGATTTGAACCTTAACCCATCAAGATCCTTAAAAATAA 642

RESULT 4
US-09-070-927A-88/c
; Sequence 88, Application us/09070927A
; Patent No. US20020120116A1

;; GENERAL INFORMATION:
;; APPLICANT: Charles A. Kunsch
;; Patrick J. Dillon
;; Steven Barash
;; TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides
;; NUMBER OF SEQUENCES: 982
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Human Genome Sciences, Inc.
;; STREET: 9410 Key West Avenue
;; CITY: Rockville
;; STATE: Maryland
;; COUNTRY: USA
;; ZIP: 20850
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
;; COMPUTER: HP Vectra 486/33
;; OPERATING SYSTEM: MSDOS version 6.2
;; SOFTWARE: ASCII Text
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/070,927A
;; FILING DATE: 04-May-2000
;; CLASSIFICATION: <Unknown>
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 60/046,655
;; FILING DATE: 1997-05-16
;; APPLICATION NUMBER: 60/044,031
;; FILING DATE: 1997-05-06
;; APPLICATION NUMBER: 60/066,009
;; FILING DATE: 1997-11-14
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Kenley K. Hoover
;; REGISTRATION NUMBER: 40,302
;; REFERENCE/DOCKET NUMBER: PB369
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (301) 309-8504
;; TELEFAX: (301) 309-8512
;; INFORMATION FOR SEQ ID NO: 88:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 6691 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; SEQUENCE DESCRIPTION: SEQ ID NO: 88:
US-09-070-927A-88

Query Match 30.1%; Score 193.2; DB 10; Length 6691;
Best Local Similarity 58.3%; Pred. No. 4.6e-47;
Matches 358; Conservative 0; Mismatches 253; Indels 3; Gaps 1;

Qy 1 ATGATTACAAGTATTTTAAATCTAGCCTATCTGCTGGTTCGATTCCTGCTGCTC 60
Db 5424 ATGAAATTCGTCATTTTGTGCTTATTTATAGTTTCGATTCCTGCTGCTGCTT 5365

Qy 61 TGGATTGGACAAGTATTTTCAATCAATCTACGCGAGCATGGTTCGTAACACTGGA 120
Db 5364 TGGATTGGACAAGTATTTTCAATCAATCTACGCGAGCATGGTTCGTAACACTGGA 5305

Qy 121 ACGACCAACACTTCCGCAATTTAGTAAAGAAAGCTGATGGCAACCTTTGTGATTCAC 180
Db 5304 ACAACCAATACATTTCTGCTTAGGGAACCTTCGCGAATACGGTATTTATATGAT 5245

Qy 181 TTTTCAAGGAACCTTAGCAGCTGCTTCCGATTTATTTTCACTACAAAGCGCTTCT 240
Db 5244 ATCTTGAAGGAACCTTAGCAGCTTCAATACCTTATTTTGTGTTTACAAGCGCTGA 5185

Qy 241 CCTCTCATCTTTGGACTTTTGGCTTTATCGGCCATACCTTCCCTATCTTTCAGGATTT 300
Db 5184 CCGTCTCTTTGGGAGTACGAGCTGTTTGGGCACTATTTTCCCTATTTTGGCAATTT 5125

Qy 301 AAAGTGTAGAGCTGTCGCAACCAAGCTGCTGGAGTATTTTCCGATTTTCCGCTATCTTC 360
Db 5124 AAAGTGTAGAGCTGTCGCAACCAAGCTGCTGGAGTATTTTCCGCTATCTTC 5065

Qy	361	TGTC	TCTACCTCGCGATTAATCTCTTTGGAGCTCTCTATCTTGGCAGTATGATTTCACTG	420
Db	5064	TTTATTATCTGCCTCTATTTTGTGTGATTTGTCGTATCTAAACACAGTATGGTGAGTTTA	5005	
Qy	421	TCTAGTGTACACAGCATCGAATCGCGCTGTATTCGGGGTTCGTCTTTCCACT---	477	
Db	5004	ACAAGTATGATTAGCGCTCTACTAATACACTTTTCTACTATTATTTGGCCTTTTCACGTGC	4945	
Qy	478	GGTTTTTATCCTGAGTAACCTATGACATCTCTCTTCATCGCTATNTATCTTAGCACTTGTCTAGT	537	
Db	4944	CCAGCTATTTTACCAACATTTTAAC7GGTATTAAACCGTGATTCGGCATTCGTTTGACACTACT	4885	
Qy	538	TTGATTATCATTTTCGTCATAAGGACAAATATAGCTCGTATCAAAAATAAAACTGAAAAATTG	597	
Db	4884	TTTATCTTTGTCGTCATCGTGAACACATTTCAGCGAATTAATAATGGACAGAAAGTCGT	4825	
Qy	598	GTCCCTTGGGGATT	611	
Db	4824	CTTTCTTTTGGCTT	4811	

RESULT 5

```

US-09-823-246-1
:
: Sequence 1, Application US/09823246
: Patent No. US20020058789A1
:
: GENERAL INFORMATION:
:
: APPLICANT: Burnham, Martin K. R.
: APPLICANT: Biswas, Sanjoy
: APPLICANT: Zalacain, Magdalena
: APPLICANT: Warren, Patrick V.
: APPLICANT: Sylvester, Daniel R.
: APPLICANT: McDevitt, Damien
:
: TITLE OF INVENTION: Ynes
:
: FILE REFERENCE: GW20001
:
: CURRENT APPLICATION NUMBER: US/09/823,246
: CURRENT FILING DATE: 2001-03-30
: PRIOR APPLICATION NUMBER: US 60/193,496
: PRIOR FILING DATE: 2000-03-31
:
: NUMBER OF SEQ ID NOS: 2
:
: SOFTWARE: FastSeq for Windows Version 4.0
:
: SEQ ID NO 1
:
: LENGTH: 609
:
: TYPE: DNA
:
: ORGANISM: Staphylococcus aureus
: US-09-823-246-1

```

Query Match 16.0%; Score 102.4; DB 10; Length 609;
Best Local Similarity 52.6%; Pred. No. 9.9e-21;
Matches 266; Conservative 0; Mismatches 216; Indels 24; Gaps 1;

Qy	1	ATGATTACAATAGTTTATTAAATCCTAGCCTATCTGCTGGTTCGATTCCATCTCGTCTC	60
Db	1	ATGATGATAATCGTGCATGTTACTACTAAGTTATCTTATCGGGCTTTCCCAAGTGGATTC	60
Qy	61	TGGATTGGCAAGTATTCTTTCAATCAATCTACGGCAGCATGGTCTCGTACACATGGA	120
Db	61	GTAATTGGCAAAATATTTTTCAAAAAAGATATTAGACAAATTTGGTAGTGGTAATACTGCG	120
Qy	121	ACGACCAACACCTTTCCGCAATTTAGGTAAAGAAAGCTGGTATGSCAACCTTTGTGATTGAC	180
Db	121	GCTACTAATAGCTTTAGAGTATTAGGTGCTGCTGCAGGATTTCTGGTAACATTTCTAGAT	180
Qy	181	TTTTTCAAAGGAACCCCTAGCAACGCTGCTTCCGATTATTTTTTCATCTCAAGGCGGTTCT	240
Db	181	ATTTTCAAAGGGTCAATACTGTTTCTTCCCTTATGGTTACAAGTTACGCAGATGGC	240
Qy	241	CCCTCTC-----ATCCTTGGACTTTTGGCTGTTATFCGGCCAT	276
Db	241	CCTATTAGTACTTTTTTACAANAATGGTTTAATTTGGTGGCTTATTTCGCTATCTTGGACAC	300
Qy	277	ACCTTCCCTATCTTGCAGAGATTTAAAGTGGTAAGGCTGTGCCAACCAAGTGCCTGGAGHG	336
Db	301	GTTTATCTCGTTTATTAAAAATTCGAAGTGGCAAGAGCTTGTGCAACTAGTGCAGGTGTC	360

RESULT 7

US-09-815-242-4250
; Sequence 4250, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A

; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US/09/815,242

; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 4250

; LENGTH: 606

; TYPE: DNA

; ORGANISM: Staphylococcus aureus

US-09-815-242-4250

Query Match 15.5%; Score 99.2; DB 10; Length 606;

Best Local Similarity 52.2%; Pred. No. 8.7e-20;

Matches 264; Conservative 0; Mismatches 218; Indels 24; Gaps 1;

QY 1 ATGATTACAAATAGTTTATTAATCTAGCTATCTGCTGGTTCGATTCATCTGCTCTC 60

DB 1 ATGATGATAATCGTCATGTACTACTAAGTTATCTATCGCGCTTTCCCAAGTGATTC 60

QY 61 TGGATTGGACAAGTATTCTTTCAATCAATCTACCGGACATGGTCTCGTAACACTGGA 120

DB 61 GTAATTGGAAAATATTATTTTCAAAAAGATATTAGACAATTTGGTAGTGGTAATACTGGC 120

QY 121 ACGACAACACCTCTCGCATTTTATAGTAAGAAAGCTGGTATGCAACCTTTCTGATTGAC 180

DB 121 GCTACTAATAGCTTTAGAGTATTAGTCTCTCGCAGGATTTCTTGGTAACATTTCTAGAT 180

QY 181 TTTTCAAGAACCCCTAGCAACGCTGCTCCGATTTATTTTCAATCAACAGCGCTTCT 240

DB 181 ATTTTCAAGGGTTCATAACTGTTTCTTCCCTTTATGTTTACCAGTTCCACGACATGC 240

QY 241 CCTCTC-----ATCTTTGGACTTTTGGCTGTTATCGGCCAT 276

DB 241 CCTATTAGTACTTTTTTACAAATGGTTTAAATTTGGTCTTATCGCTACTATGACAC 300

QY 277 ACCTTCCCTATCTTTCAGGATTTAAAGTGTGAAGGCTGTCGCAACAGTCTCGAGTG 336

DB 301 GTTTATCTGTTTATTTAAATTCAGGTGGCAAGCGGTTCGAACACTAGTCAGGTGC 360

QY 337 ATTTTCGGATTTGCGCTATCTTCTCTACTACCTTCGATTTATCTTTTGGAGCTCTC 396

DB 361 GTCTTGGAGTCAACCCGACTACTTTTACTAATACTTTGCAATTTACTCTTTATTGATG 420

QY 397 TATCTTGGCAGTATGATTTCACCTAGTGTACAGCATCGATTGGCGCTGTTATCGG 456

DB 421 AAGATTTTAAATATGTTTCTTACCAAGTATCGTTGAGCAATTTGCTGTGATGCGC 480

QY 457 GTCTGCTCTTTTCCACTTTTGGTTT 482

DB 481 TCGCTTATCATTCATCAAGACTATATTTT 506

RESULT 8

US-09-815-242-8039

; Sequence 8039, Application US/09815242

; Patent No. US20020061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in

; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242

; CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 8039

; LENGTH: 609

; TYPE: DNA

; ORGANISM: Staphylococcus aureus

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)...(609)

US-09-815-242-8039

Query Match 15.5%; Score 99.2; DB 10; Length 609;

Best Local Similarity 52.2%; Pred. No. 8.7e-20;

Matches 264; Conservative 0; Mismatches 218; Indels 24; Gaps 1;

QY 1 ATGATTACAAATAGTTTATTAATCTAGCTATCTGCTGGTTCGATTCATCTGCTCTC 60

DB 1 ATGATGATAATCGTCATGTACTACTAAGTTATCTATCGCGCTTTCCCAAGTGATTC 60

QY 61 TGGATTGGACAAGTATTCTTTCAATCAATCTACCGGACATGGTCTCGTAACACTGGA 120

DB 61 GTAATTGGAAAATATTATTTTCAAAAAGATATTAGACAATTTGGTAGTGGTAATACTGGC 120

QY 121 ACGACAACACCTCTCGCATTTTATAGTAAGAAAGCTGGTATGCAACCTTTCTGATTGAC 180

DB 121 GCTACTAATAGCTTTAGAGTATTAGTCTCTCGCAGGATTTCTTGGTAACATTTCTAGAT 180

QY 181 TTTTCAAGAACCCCTAGCAACGCTGCTCCGATTTATTTTCAATCAACAGCGCTTCT 240

DB 181 ATTTTCAAGGGTTCATAACTGTTTCTTCCCTTTATGTTTACCAGTTCCACGACATGC 240

Qy	241	CCCTCTC-----ATCTTTGGACTTTTGGCTGTTATCGGCCAT	276
Db	241	CCATATTAGTACTTTTTTACAAATGGTTTAAATTTTGGCTTATCTGCTATCTATTGGACAC	300
Qy	277	ACCTTCCTCATCTTTGCGAGATTTAAAGTGGTAAGCTCTGCCAACCACTGCTGGAGTGC	336
Db	301	GTTTATCTCTTATTAAATTTCCAAAGTGGCAAGCGTTGCACTAGTCGAGTGTTC	360
Qy	337	ATTTTCGGATTTGGGCTCATCTCTCTGTCCTACCTTCGCAATATCTTCTTTGGAGCTCTC	396
Db	361	GTCTTGGGAGTCAACCGGATCTTTTACTAATACTTGTCAATTATCTTCTTATTGTATTG	420
Qy	397	TATCTTGGCAGTAGTATTTCCACTGCTAGTGTACACAGATCGATTGGCGCTGTTATTCGGG	456
Db	421	AAGATTTTTAAATATGTTTCTTTAGCAAGTATCTTTCGACAAATTTGCTGTGTGATTGGC	480
Qy	457	GTCTGCTCTTTCCACTTTTGGTTT	482
Db	481	TCGCTTATCATTTCAAGACTATATTTT	506

RESULT 9

US-09-815-242-6900
 : Sequence 6900, Application US/09815242
 : Patent No. US20020061569A1
 : GENERAL INFORMATION:
 : APPLICANT: Haselbeck, Robert
 : APPLICANT: Ohlsen, Kari L.
 : APPLICANT: Zyskind, Judith W.
 : APPLICANT: Wall, Daniel
 : APPLICANT: Trawick, John D.
 : APPLICANT: Carr, Grant J.
 : APPLICANT: Yamamoto, Robert T.
 : APPLICANT: Xu, H. Howard

```

1 TITLE OF INVENTION: Identification of Essential Genes in
2
3 TITLE OF INVENTION: Prokaryotes
4
5 FILE REFERENCE: ELITRA.011A
6
7 CURRENT APPLICATION NUMBER: US/09/815,242
8
9 CURRENT FILING DATE: 2001-03-21
10
11 PRIOR APPLICATION NUMBER: 60/191,078
12
13 PRIOR FILING DATE: 2000-03-21
14
15 PRIOR APPLICATION NUMBER: 60/206,848
16
17 PRIOR FILING DATE: 2000-05-23
18
19 PRIOR APPLICATION NUMBER: 60/207,727
20
21 PRIOR FILING DATE: 2000-05-26
22
23 PRIOR APPLICATION NUMBER: 60/242,578
24
25 PRIOR FILING DATE: 2000-10-23
26
27 PRIOR APPLICATION NUMBER: 60/253,625
28
29 PRIOR FILING DATE: 2000-11-27
30
31 PRIOR APPLICATION NUMBER: 60/257,931
32
33 PRIOR FILING DATE: 2000-12-22
34
35 PRIOR APPLICATION NUMBER: 60/269,308
36
37 PRIOR FILING DATE: 2001-02-16
38
39 NUMBER OF SEQ ID NOS: 14110
40
41 SOFTWARE: FastSeq for Windows Version 4.0

```

Query Match 10.5%; Score 67.4; DB 10; Length 600;
Best Local Similarity 50.5%; Pred. No. 2.2e-10;
Matches 164; Conservative 0; Mismatches 161; Indels 0; Gaps 0;

Qy 14 TTTTATTAATCCTAGCCTATCTGCTGGGTTCCGATTCGATCTGGTCTCTGGATTGGACAAG 73
 Db 20 TTTATATGCTTTTTCCTATCTTTTAGGCTCAATTTCCAGTCGATTTTAAATTTCTCGCA 79
 Qy 74 TATCTCTTCAANTCAATCTACCGCAGCATGGTTCTGGTAACTGGAACGACCAACACT 133

Db	80	TTGGAGGCTTCCAGATCCTCGTCAAAATGGTTCACACACCTGGGGGACAAATGCT	139
Qy	134	TCCGCATTTTAGTAAAGACGTGGTATGGCAACCTTTTGATGACTTTTTCAAAGGA	193
Db	140	TCCGCATTTGGAATCGTAACTCGCCTGGCGTGCTATTTTGTATGTTTAAAGGA	199
Qy	194	CCCTTAGCAACGCTGCTTCCGATTATTTTTCATCTACAAGCGTTTCTCCTCTCATCTTG	253
Db	200	TGATTCTGTATGGCAGGCTATTATTATTAGTTTTAACCCAATTTGAATGGGAATG	259
Qy	254	GACTTTTGGCTGTATCGGCCATCCTTCCCTCATCTTGCAGGATTTAAAGTGGTAAGG	313
Db	260	PCCTTGGTGCTTGTTTAGGCGCATATTTTCCCAATTTTCTTTCAAATTAAGGTGGCAAG	319
Qy	314	CTGTCCCAACGAGTCTGGAGTGAT	338
Db	320	GTGTTGCAACTGCCCTTTGGGCAAT	344

RESULT 10

```

US-09-815-242-7702
: Sequence 7702, Application US/09815242
: Patent No. US20020061569A1
: GENERAL INFORMATION:
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlsen, Kari L.
: APPLICANT: Zykind, Judith W.
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John D.
: APPLICANT: Carr, Grant J.
: APPLICANT: Yamamoto, Robert T.
: APPLICANT: Xu, H. Howard
: TITLE OF INVENTION: Identification of Ess
: TITLE OF INVENTION: Prokaryotes
: FILE REFERENCE: ELITRA.011A
: CURRENT APPLICATION NUMBER: US/09/815,242
: CURRENT FILING DATE: 2001-03-21
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/242,578
: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: NUMBER OF SEQ ID NOS: 14110
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 7702
: LENGTH: 570
: TYPE: DNA
: ORGANISM: Pseudomonas aeruginosa
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (1)...(570)
US-09-815-242-7702

```

Query Match	9.2%	Score 59;	DB 10;	Length 570;
Best Local Similarity	48.9%	Pred.	NO.	6.4e-08;

QY	22	ATCCTAGCCTATCTGCTGGGTTCGATTCCACTCTGGTCTTGGATTGGACAAGTATCTTTT	81
Db	19	ATCCTCGCCTACCTGCTCGGTTCCTCTCCCTTCGCCGTCCTGCTCAGCCGCTGTTCCGC	78
QY	82	CAAAATCAATCTACCGGAGCATGTTCTTGGTAACACTGGAAGACCAACACCTTCGCGATT	147

Db 79 ACCAGGACCGCGCGCGGCTTCGGCAACCCCGCGGACCAACATGCTGCGGCTC 138
Qy 142 TTAGTAAGAAGCTGGTATGCAACCTTTGTGATTGACTTTTTCARAGGAACCTAGCA 201
Db 139 GCGGGAAGAACTCGCCATCTGACCTGCTCGCGACGTCGGCAAGGCGCTGTGCGG 198
Qy 202 AGCTGCTTCGGAATATTTTTCATCTACAAGCGGTTTCTCTCTCATCTTTGGACTTTTG 261
Db 199 GTGCTGGTCGCCGCTGGCTCGCGTGATGAGGAGGCTGGGTGCGGATCGCC 258
Qy 262 GCTGTATCGGCCATACCTTCCTATCTTTGCAAGATTTAAAGGTGTAAGCTGTCCGA 321
Db 259 GCGGTGATCGGCGACCTGTACCGCTGACTTCAACTTCCGCGCGCAAGGGTGTGCGC 318
Qy 322 ACCAGTGTGGAGTGAATTTTCG 344
Db 319 ACCGCGCGGCGCATGCTCTCGG 341

RESULT 11

US-09-815-242-9659

; Sequence 9659, Application US/09815242

; Patent No. US20020061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes In

; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242

; CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 9659

; LENGTH: 612

; TYPE: DNA

; ORGANISM: Salmonella typhi

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1)...(612)

US-09-815-242-9659

Query Match 9.0%; Score 57.8; DB 10; Length 612;
Best Local Similarity 48.9%; Pred. No. 1.5e-07;
Matches 155; Conservative 0; Mismatches 162; Indels 0; Gaps 0;

Qy 23 TCCTAGCCATCTCGTGGTTCGATTCCTGCTGCGGATGCGAAGTATCTTTC 82
Db 29 TCCTCGGTCACCTCTCGGCTCAATTTCCAGTCCCATCTGTTTCCCGCATCGCCGGC 88
Qy 83 AAATCAATCTACCGGACGATGGTTCTGGTAACACTGGAACGACCAACCTTCGCGCATTT 142
Db 89 TGCCCGATCGCGAGAAAGCGGCTCCGGCAACCCCGCGGCGGAGGATTTACGAATCG 148

Qy 143 TAGTAAGAAGCTGGTATGCAACCTTTGTGATTGACTTTTTCAAAGGAACCTAGCAA 202
Db 149 GTGCAAGGAGCGCTCTCGCGTACTTATTTTGGACATCTTAAAGGCATGTTGCCG 208
Qy 203 CGCTGCTTCCGATATTTTTCATCTACAAGCGGTTTCTCTCTCATCTTTGGACTTTTG 262
Db 209 TCTGGGCGCGTATGCGTTCAGCGTTACCCCTTCTTGCTGGGCTTATCGCTATCGCG 268
Qy 263 CTGTATCGGCCATACCTTCCCTATCTTTGCGAGGATTTTAAAGGTGTAAGCTGTGCGCA 322
Db 269 CTTGCTGGGACATCTTGGCGGCTCTTTTGGCTTTAAAGGCGGGAAGCGGTAGCAA 328
Qy 323 CCAGTGTCTGGAGTGATT 339
Db 329 CCGCGTTTGGCGCCATT 345

RESULT 12

US-09-815-242-3464/c

; Sequence 3464, Application US/09815242

; Patent No. US20020061569A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Ohlsen, Kari L.

; APPLICANT: Zyskind, Judith W.

; APPLICANT: Wall, Daniel

; APPLICANT: Trawick, John D.

; APPLICANT: Carr, Grant J.

; APPLICANT: Yamamoto, Robert T.

; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes In

; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242

; CURRENT FILING DATE: 2001-03-21

; PRIOR APPLICATION NUMBER: 60/191,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 3464

; LENGTH: 189

; TYPE: DNA

; ORGANISM: Staphylococcus aureus

US-09-815-242-3464

Query Match 8.8%; Score 56.6; DB 10; Length 189;
Best Local Similarity 62.2%; Pred. No. 1.9e-07;
Matches 89; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

Qy 243 TCCATCTTTGGACTTTTGGCTGTTATCGGCCATACCTTCCCTATCTTTGCGAGGATTTAA 302
Db 143 TTAAATTTGTTGCTTATTCGCTATCTATCTGACACGTTTATCTGTTTATTTAAATTTCA 84
Qy 303 AGCTGTAAGGCTGTCGCAACAGTGTGAGTGTATTTTCGGATTTGCCCTATCTTCG 362
Db 83 AGCTGCAAGCTGTTGCAACTAGTGCAGTGTCTGCTTGGGAGTCAATCCGATCTTTT 24
Qy 363 TCTCTACCTTGGGATTTATCTTCT 385
Db 23 ACTAATACITGGCAATTTATCTTCT 1

Search completed: December 26, 2002, 00:32:36
Job time : 194.61 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 25, 2002, 14:17:36 ; Search time 468.714 Seconds
(without alignments)
54.868 Million cell updates/sec

Title: US-10-068-080-3
Perfect score: 960
Sequence: 1 MLIALIILAYLIGSIPGL.....RHRANIKRIINKTEPKVKWL 193

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : 1: A_Geneseq_101002.*
1: /SID22/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID22/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	960	100.0	193	20	AA1980
2	960	100.0	193	21	AA1981
3	566.5	59.0	198	23	AB48775
4	476.5	49.6	204	23	AB40394
5	468	48.8	213	21	AA192246
6	468	48.8	213	22	AAU37736
7	468	48.8	213	22	AAU37943
8	465.5	48.5	197	22	AAU32881
9	465.5	48.5	202	22	AAU33809
10	465.5	48.5	202	22	AAU36544

11	454	47.3	212	23	ABP26818
12	453	47.2	240	23	ABP26819
13	447	46.6	207	23	ABP30349
14	445	46.4	213	23	ABP30349
15	275.5	28.7	114	20	AA1922561
16	261.5	27.2	205	22	AAU36120
17	257.5	26.8	203	22	AAU38164
18	255	26.6	199	22	AAU35404
19	252	26.2	189	22	AAU36206
20	250.5	26.1	205	22	AAU34710
21	230	24.0	220	22	AAU36021
22	228	23.8	262	19	AAW98757
23	228	23.8	262	22	AAU35860
24	226	23.5	173	21	AAU74918
25	226	23.5	173	21	AAU74919
26	224	23.3	173	21	AAU74917
27	94.5	9.8	316	22	AAU44251
28	94	9.8	981	20	AAU34476
29	94	9.8	985	20	AAU34352
30	93.5	9.7	299	22	AB887731
31	93	9.7	389	23	AB853723
32	93	9.7	489	23	ABP40396
33	92.5	9.6	451	22	AAU34954
34	92.5	9.6	1014	22	AAU34954
35	92	9.6	336	23	ABP26867
36	91	9.5	783	22	ABP79006
37	91	9.5	801	22	AAU30006
38	89.5	9.3	451	23	ABP29492
39	89.5	9.3	463	22	AAU31628
40	89	9.3	155	21	AAU51973
41	89	9.3	188	21	AAU51965
42	89	9.3	251	21	AAU51964
43	88.5	9.2	393	23	AB853458
44	88	9.2	248	22	AAU31769
45	88	9.2	254	23	AAU75790

ALIGNMENTS

RESULT 1
AA1980
ID AA1980 standard; Protein: 193 AA.
AC AA1980;
XX AA1980;
XX 17-NOV-1999 (first entry)
XX B. subtilis B-ynes protein sequence.
XX General essential protein; pathogenic bacteria; pathogen; inhibitor;
XX Bacterial growth; B-ynes.
XX Bacillus subtilis.
XX WO9933871-A2.
XX 08-JUL-1999.
XX 30-DEC-1998; 98WO-US27918.
XX 31-DEC-1997; 97US-0070116.
XX (MILL-) MILLENNIUM PHARM INC.
XX Youngman P, Fritz C, Murphy C, Guzman L;
XX WPI; 1999-430230/36.
XX N-PSDB; AA20371.
XX Streptococcus pneumoniae general essential protein genes and proteins,
XX useful for identification of antibacterial agents.

Streptococcus poly	11	454	47.3	212	23	ABP26818
Streptococcus poly	12	453	47.2	240	23	ABP26819
Streptococcus poly	13	447	46.6	207	23	ABP30349
Lactococcus lactis	14	445	46.4	213	23	ABP30349
Bacterial general	15	275.5	28.7	114	20	AA1922561
Klebsiella pneumoniae	16	261.5	27.2	205	22	AAU36120
Salmonella typhi c	17	257.5	26.8	203	22	AAU38164
Haemophilus influenzae	18	255	26.6	199	22	AAU35404
Pseudomonas aeruginosa	19	252	26.2	189	22	AAU36206
E. coli cellular p	20	250.5	26.1	205	22	AAU34710
Helicobacter pylori	21	230	24.0	220	22	AAU36021
H. pylori GHP108	22	228	23.8	262	19	AAW98757
Helicobacter pylori	23	228	23.8	262	22	AAU35860
Neisseria meningitidis	24	226	23.5	173	21	AAU74918
Neisseria meningitidis	25	226	23.5	173	21	AAU74919
Neisseria gonorrhoea	26	224	23.3	173	21	AAU74917
Propionibacterium	27	94.5	9.8	316	22	AAU44251
Porphorymonas ging	28	94	9.8	981	20	AAU34476
Porphorymonas ging	29	94	9.8	985	20	AAU34352
Human T2R01 amino	30	93.5	9.7	299	22	AB887731
Lactococcus lactis	31	93	9.7	389	23	AB853723
Staphylococcus epi	32	93	9.7	489	23	ABP40396
Enterococcus faeca	33	92.5	9.6	451	22	AAU34954
C glutamicum prote	34	92.5	9.6	1014	22	AAU34954
Streptococcus poly	35	92	9.6	336	23	ABP26867
C. glutamicum SRT	36	91	9.5	783	22	ABP79006
C. glutamicum prote	37	91	9.5	801	22	AAU30006
Streptococcus poly	38	89.5	9.3	451	23	ABP29492
S. epidermidis ope	39	89.5	9.3	463	22	AAU31628
Arabidopsis thalia	40	89	9.3	155	21	AAU51973
Arabidopsis thalia	41	89	9.3	188	21	AAU51965
Arabidopsis thalia	42	89	9.3	251	21	AAU51964
Streptococcus lactis	43	88.5	9.2	393	23	AB853458
S. epidermidis ope	44	88	9.2	248	22	AAU31769
Human protein phos	45	88	9.2	254	23	AAU75790

PS Disclosure; Fig 24; 124pp; English.
XX This sequence is the Bacillus subtilis B-ynes protein. B-ynes is
CC related to the Streptococcus pneumoniae general essential
CC protein (GEP) gene of the invention. The genes encoding the GEP
CC polypeptides are useful molecular tools for identifying similar genes in
CC pathogenic microorganisms, such as pathogenic strains of Bacillus. In
CC addition, the operons containing genes encoding GEP and the polypeptides
CC themselves, are useful targets for identifying compounds that are
CC inhibitors of the pathogens in which the GEP are expressed. Such
CC inhibitors are useful for inhibiting bacterial growth by being
CC bacteriostatic or bacteriocidal.

XX Sequence 193 AA;
SQ Query Match 100.0%; Score 960; DB 20; Length 193;
Best Local Similarity 100.0%; Pred. No. 9e-100;
Matches 193; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MLIALLIILAYLIGSPGLIVGKLAKGIDIREHSGNIGATNAFTLGVKAGSVVIAGD 60
DB 1 MLIALLIILAYLIGSPGLIVGKLAKGIDIREHSGNIGATNAFTLGVKAGSVVIAGD 60
QY 61 ILKGTALATLAPFLMHVDIHPHLLAGVFAVLGHVPPPIFAKFKGKAVATSGVLLFYAPLLF 120
DB 61 ILKGTALATLAPFLMHVDIHPHLLAGVFAVLGHVPPPIFAKFKGKAVATSGVLLFYAPLLF 120
QY 121 ITWVAVFFIPLYTKFVLSLSSMLTGIVTYISFFVHDTYLLIVVTLTIFVIYRHRANIK 180
DB 121 ITWVAVFFIPLYTKFVLSLSSMLTGIVTYISFFVHDTYLLIVVTLTIFVIYRHRANIK 180
QY 181 RIINKTEPKVKWL 193
DB 181 RIINKTEPKVKWL 193

RESULT 2
AAY92247
ID AAY92247 standard; Protein; 193 AA.
XX AC AAY92247;
XX DT 10-AUG-2000 (first entry)
XX DE B. subtilis B-ynes polypeptide.
XX KW B-ynes; survival; antibacterial; inhibitor.
XX OS Bacillus subtilis.
XX PN WO200020627-A1.
XX PD 13-APR-2000.
XX PF 30-SEP-1999; 99WO-US22665.
XX PR 30-SEP-1998; 98US-0163445.
XX PA (MILL-) MILLENNIUM PHARM INC.
XX PI Fritz C, Youngman P, Guzman L;
XX WPI; 2000-303799/26.
XX DR N-PSDB: AAA09182.
XX PT Methods for identifying an antibacterial agent for treating
XX Streptococcus pneumoniae infections comprises detecting an interaction
XX between a yneS polypeptide and a test compound
XX Disclosure; Fig 2; 65pp; English.
XX Bacillus subtilis yneS gene is a homologue of Streptococcus pneumoniae
XX yneS gene, which encodes a polypeptide (S-ynes) essential for survival

CC for a wide range of bacteria. Identifying an antibacterial agent
CC comprises contacting a yneS polypeptide (S-ynes) with a test compound and
CC detecting an interaction of the test compound with the S-ynes polypeptide
CC which indicates that the compound is an antibacterial agent.
CC Alternatively, detecting a decrease in function of the polypeptide
CC contacted with the test compound and determining whether the compound
CC inhibits growth of bacteria, relative to the growth of bacteria cultured
CC in the absence of a test compound where inhibition of growth indicates
CC the compound is an antibacterial agent. Inhibitors of S-ynes function
CC are useful for treating a Streptococcus pneumoniae infection in mammals.
XX Sequence 193 AA;

XX Query Match 100.0%; Score 960; DB 21; Length 193;
Best Local Similarity 100.0%; Pred. No. 9e-100;
Matches 193; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MLJALLIILAYLIGSPGLIVGKLAKGIDIREHSGNIGATNAFTLGVKAGSVVIAGD 60
DB 1 MLJALLIILAYLIGSPGLIVGKLAKGIDIREHSGNIGATNAFTLGVKAGSVVIAGD 60
QY 61 ILKGTALATLAPFLMHVDIHPHLLAGVFAVLGHVPPPIFAKFKGKAVATSGVLLFYAPLLF 120
DB 61 ILKGTALATLAPFLMHVDIHPHLLAGVFAVLGHVPPPIFAKFKGKAVATSGVLLFYAPLLF 120
QY 121 ITWVAVFFIPLYTKFVLSLSSMLTGIVTYISFFVHDTYLLIVVTLTIFVIYRHRANIK 180
DB 121 ITWVAVFFIPLYTKFVLSLSSMLTGIVTYISFFVHDTYLLIVVTLTIFVIYRHRANIK 180
QY 181 RIINKTEPKVKWL 193
DB 181 RIINKTEPKVKWL 193

RESULT 3
ABB48775
ID ABB48775 standard; Protein; 198 AA.
XX AC ABB48775;
XX DT 05-FEB-2002 (first entry)
XX DE Listeria monocytogenes protein #1479.
XX KW Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
XX vitamin B12; bacterial infection; disease.
XX OS Listeria monocytogenes.
XX PN WO200177335-A2.
XX PD 18-OCT-2001.
XX PF 11-APR-2001; 2001WO-FR01118.
XX PR 11-APR-2000; 2000FR-0004629.
XX PA (INSP) INST PASTEUR.
XX Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;
XX Dussurget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P;
XX Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;
XX Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
XX Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durand L;
XX Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
XX Maduenio E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;
XX Rose M, Voss H;
XX WPI; 2002-010914/01.
XX Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
XX and prevention of Listeria and related bacterial infections, and
XX related polypeptides

PT Streptococcus pneumoniae infections comprises detecting an interaction
 PT between a ynes polypeptide and a test compound
 PS Disclosure: Fig 1: 65pp: English.
 XX
 CC Streptococcus pneumoniae ynes gene encodes a polypeptide (S-ynes)
 CC essential for survival for a wide range of bacteria. Identifying an
 CC antibacterial agent comprises contacting a ynes polypeptide (S-ynes) with
 CC a test compound and detecting an interaction of the test compound with
 CC the S-ynes polypeptide which indicates that the compound is an
 CC antibacterial agent. Alternatively, detecting a decrease in function of
 CC the polypeptide contacted with the test compound and determining whether
 CC the compound inhibits growth of bacteria, relative to the growth of
 CC bacteria cultured in the absence of a test compound where inhibition of
 CC growth indicates the compound is an antibacterial agent. Inhibitors of
 CC S-ynes function are useful for treating a Streptococcus pneumoniae
 CC infection in mammals.
 XX
 SQ Sequence 213 AA;
 Query Match 48.8%; Score 468; DB 21; Length 213;
 Best Local Similarity 48.0%; Pred. No. 2.1e-44;
 Matches 97; Conservative 32; Mismatches 63; Indels 10; Gaps 3;
 QY 1 MLIALLLIAYLGISGLIVGKLGKIDIREHSGNLGATNAPRTILGVKAGSVVIAGD 60
 DB 1 MITIVLLIAYLGISGLIVGKLGKIDIREHSGNLGATNAPRTILGVKAGMATFVID 60
 QY 61 ILKGTALATPLMHVD-IHPLLAGVFAVLGHVPIFAKFGKAVATSGGVLLFYAPLL 119
 DB 61 FFKGTALATPLPIIFHLQGVSPILFGLLAVIGHTTPIFAGFGKAVATSGVIFGPAPIF 120
 QY 120 FITWAVFFIYLYTKFVLSMLTGIVTVI-----YSFFV--HDTYLLIVVTLTIF 170
 DB 121 CLYLAIFFGALYLGSMISLSSVTASIAAVIGVLLPFLFGFILSNYDSLFATIALASL 180
 QY 171 VIYHRANKRIINKTEPKVKW 192
 DB 181 IIRHKDNIARIKNTENLVPM 202
 RESULT 6
 AAU37736
 ID AAU37736 standard; Protein: 213 AA.
 AC AAU37736;
 XX
 DT 14-FEB-2002 (first entry)
 DE Streptococcus pneumoniae cellular proliferation protein #165.
 KW Antisense; prokaryotic cellular proliferation protein;
 KW antibiotic; antibacterial; drug design.
 OS Streptococcus pneumoniae.
 XX WO200170955-A2.
 XX 27-SEP-2001.
 PD
 PF 21-MAR-2001; 2001WO-US09180.
 PR 21-MAR-2000; 2000US-191078P.
 PR 23-MAY-2000; 2000US-206948P.
 PR 26-MAY-2000; 2000US-207727P.
 PR 23-OCT-2000; 2000US-242578P.
 PR 27-NOV-2000; 2000US-253625P.
 PR 22-DEC-2000; 2000US-257931P.
 PR 16-FEB-2001; 2001US-269308P.
 XX
 XX (ELIT-) ELITRA PHARM INC.
 XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
 PI

PI Yamamoto RT, Xu HH;
 XX
 DR WPI: 2001-611495/70.
 DR N-PSDB: AAS55595.
 XX
 PT New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids -
 XX
 PS Example 3; Seq ID No 13329; 51pp; English.
 CC
 CC The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence represents an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 213 AA;
 Query Match 48.8%; Score 468; DB 22; Length 213;
 Best Local Similarity 48.0%; Pred. No. 2.1e-44;
 Matches 97; Conservative 32; Mismatches 63; Indels 10; Gaps 3;
 QY 1 MLIALLLIAYLGISGLIVGKLGKIDIREHSGNLGATNAPRTILGVKAGSVVIAGD 60
 DB 1 MITIVLLIAYLGISGLIVGKLGKIDIREHSGNLGATNAPRTILGVKAGMATFVID 60
 QY 61 ILKGTALATPLMHVD-IHPLLAGVFAVLGHVPIFAKFGKAVATSGGVLLFYAPLL 119
 DB 61 FFKGTALATPLPIIFHLQGVSPILFGLLAVIGHTTPIFAGFGKAVATSGVIFGPAPIF 120
 QY 120 FITWAVFFIYLYTKFVLSMLTGIVTVI-----YSFFV--HDTYLLIVVTLTIF 170
 DB 121 CLYLAIFFGALYLGSMISLSSVTASIAAVIGVLLPFLFGFILSNYDSLFATIALASL 180
 QY 171 VIYHRANKRIINKTEPKVKW 192
 DB 181 IIRHKDNIARIKNTENLVPM 202
 RESULT 7
 AAU37943
 ID AAU37943 standard; Protein: 213 AA.
 AC AAU37943;
 XX
 DT 14-FEB-2002 (first entry)
 DE Streptococcus pneumoniae cellular proliferation protein #372.
 KW Antisense; prokaryotic cellular proliferation protein;
 KW antibiotic; antibacterial; drug design.
 OS Streptococcus pneumoniae.
 XX WO200170955-A2.
 XX 27-SEP-2001.
 PD
 PF 21-MAR-2001; 2001WO-US09180.

```
XX 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
XX WPI: 2001-611495/70.
DR N-PSDB: AAS55802.
XX
XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
XX Example 3; Seq ID No 13536; 511pp; English.
XX
XX The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 213 AA;
Query Match 48.8%; Score 468; DB 22; Length 213;
Best Local Similarity 48.0%; Pred. No. 2.1e-44;
Matches 97; Conservative 32; Mismatches 63; Indels 10; Gaps 3;
QY 1 MLTALLTILAYLIGSPGLVGLKAGIDIREHSGNGLGATNAPRTGLGKAGSVVIAGD 60
DB 1 MTIVLLILLAYLLGSPGLWIGQVFFQINLREHSGNGTNTFRILCKKAGMATFVID 60
QY 61 ILKGTALATLPLMHVD-IHPLLACGFVAVLGHVFPFIFAKFGKGAATSGGVLLFYAPLL 119
DB 61 PFKGTLATLPIIFHLQGVSPILFGLLAVIGHTPTPIFAGFGKGAATSGAGVIFGAPFI 120
QY 120 FITWAVVFIFLYLTKFVLSLSMLTGIVTI-----YSFV--HDYLLIVVTLIRIF 170
DB 121 CLYLAIIFFGALYLSMSLSSTASIAAVIGLVLLFPLFGFTLSNYDSLFTAILALASL 180
QY 171 VIYHRANKRIINKTEPKVKW 192
DB 181 IIRHKDNIARIKNKTENLVPW 202
XX
XX AAG82881
ID AAG82881 standard; Protein; 197 AA.
XX
XX AAG82881;
XX
XX 03-SEP-2001 (first entry)
XX
```

```
DE S. epidermidis open reading frame protein sequence SEQ ID NO:2856.
XX
XX Staphylococcus epidermidis SRI strain; infection; diagnosis;
KW vaccination; endocarditis.
XX
XX Staphylococcus epidermidis.
XX
XX WO200134809-A2.
XX
XX 17-MAY-2001.
XX
XX 09-NOV-2000; 2000WO-US30782.
XX
XX 09-NOV-1999; 99US-0164258.
XX
XX (GLAX ) GLAXO GROUP LTD.
XX
XX Kimmerly WJ;
XX
XX WPI: 2001-316495/33.
DR N-PSDB: AAH53731.
XX
XX Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
PT useful for vaccinating against infections, e.g. endocarditis -
XX
XX Claim 13; Page 745; 2188pp; English.
XX
XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
CC (I) and (II) can have antibacterial activity and therefore can be used
CC in vaccination. The nucleic acids (I) may be used to produce the
CC S. epidermidis polypeptides (II) via the production of vectors
CC containing them which are used to produce hosts cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to
CC AAH55098 represent oligonucleotide sequences and primers which are used
CC in the exemplification of the present invention.
CC N.B. The present invention specifically claims all the polynucleotide
CC sequences given in the sequence listing of the present specification,
CC however the sequence listing only goes up to SEQ ID NO:4454 so even,
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
CC no sequences are present for SEQ ID NO:4455 to 4464.
XX
XX Sequence 197 AA;
```

```
Query Match 48.5%; Score 465.5; DB 22; Length 197;
Best Local Similarity 48.7%; Pred. No. 3.7e-44;
Matches 96; Conservative 35; Mismatches 57; Indels 9; Gaps 2;
QY 6 LTIILAYLIGSPGLVGLKAGIDIREHSGNGLGATNAPRTGLGKAGSVVIAGDILKGT 65
DB 1 MLILSYLCAPFSGLLIGLCKLFKKDIRGYSGNTGATNSFRVLGRPAGFIVTFPIKGF 60
QY 66 LATALP--FLMHVD-----IHPLLACGFVAVLGHVFPFIFAKFGKGAATSGGVLLFYA 116
DB 61 ITVFPLWFPVHADVISTFTFTNGLVGLFALGHVVPYILKFKNGKGAATSGAVLGVN 120
QY 117 PLLFITWAVVFIFLYLTKFVLSLSMLTGIVTVIYVSFFVHDYLLIVVTLITFIVYHR 176
DB 121 PILLLILAIIFFSVLKIFKYVLSLSIIAAISCVIGSIHDIHYLLAVSGIVSIIILIRHK 180
QY 177 ANIKRIINKTEPKVKWL 193
DB 181 SNIVRIFKGEPEPKKNW 197
```

RESULT 9
AAU33809

AAU33809 standard; Protein; 202 AA.
AAU33809;
14-FEB-2002 (first entry)
Staphylococcus aureus cellular proliferation protein #85.
Antisense; prokaryotic cellular proliferation protein;
antibiotic; antibacterial; drug design.
Staphylococcus aureus.
WO200170955-A2.
27-SEP-2001.
21-MAR-2001; 2001WO-US09180.
21-MAR-2000; 2000US-191078P.
23-MAY-2000; 2000US-206848P.
26-MAY-2000; 2000US-207727P.
23-OCT-2000; 2000US-242578P.
27-NOV-2000; 2000US-253625P.
22-DEC-2000; 2000US-257931P.
16-FEB-2001; 2001US-269308P.
(ELIT-) ELITRA PHARM INC.
Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
Yamamoto RT, Xu HH;
WPI: 2001-611495/70.
N-PSDB; AAS51668.
New polynucleotides for the identification and development of
antibiotics, comprise sequences of antisense nucleic acids -
Example 3; Seq ID No 5305; 511pp; English.
The invention relates to antisense inhibitors of genes essential to
prokaryotic cellular proliferation, their use in identifying the
genes, their use in the discovery of novel antibiotics, the essential
genes themselves and the encoded proteins. The prokaryotes used are
Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
invention is also useful for the identification of potential new targets
to identify proteins used in proliferation, to express these proteins,
and to obtain antibodies capable of binding to the expressed proteins.
The proteins can be used to screen compounds in rational drug discovery
programmes. The antisense nucleic acid sequence is also useful to screen
for homologous nucleic acids which are required for cell proliferation in
essential prokaryotic cellular proliferation protein.
Note: The sequence data for this patent did not form part
of the printed specification, but was obtained in electronic
format directly from WIPO at
ftp.wipo.int/pub/published_pct_sequences.

Query Match 48.5%; Score 465.5; DB 22; Length 202;
Best Local Similarity 46.5%; Pred. No. 3.8e-44;
Matches 94; Conservative 40; Mismatches 59; Indels 9; Gaps 2;
OY 1 MLIALLIILAYIGSPGLVIGKLAGIDIREHSGNLTGATNAPRTLCVKGAGSVVIAGD 60
DB 1 MMIIWLLSYLIGFSPGVLGKLFKKDKIQFGSGNTGATNSFRVLCRAGFLVTFID 60
OY 61 ILKGTALATPFLM--HVD-----IHPLLAGVFAVLGHVFPFPAKFGKAVATSGV 111
DB 61 IFKGFITVFPFLPVLPHADGPISTFTFTNGLIIVGLFAILGHVYVYLVKFGQCKAVATSGV 120

OY 112 LLEYAPLLFITMVAFFIFLYLTKEVSLSSMLTGITYTVIYFFVHDYVLLIIVTLLTIFV 171
DB 121 VLGVPNPIILLILALIIFFIVLKIFKYVSLASIVAACCIGSLTIQDYILLVWSFLVSL 180
OY 172 IYVHRANIKRIINTEPKVKWL 193
DB 181 IIRHSNISRIFRGEEPAIKWM 202
RESULT 10
AAU36544
ID AAU36544 standard; Protein; 202 AA.
AC AAU36544;
DT 14-FEB-2002 (first entry)
DE Staphylococcus aureus cellular proliferation protein #714.
KW Antisense; prokaryotic cellular proliferation protein;
antibiotic; antibacterial; drug design.
OS Staphylococcus aureus.
PN WO200170955-A2.
PD 27-SEP-2001.
PF 21-MAR-2001; 2001WO-US09180.
PR 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
PA (ELIT-) ELITRA PHARM INC.
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX WPI: 2001-611495/70.
XX N-PSDB; AAS54403.
XX New polynucleotides for the identification and development of
XX antibiotics, comprise sequences of antisense nucleic acids -
XX Example 3; Seq ID No 12137; 511pp; English.
XX The invention relates to antisense inhibitors of genes essential to
XX prokaryotic cellular proliferation, their use in identifying the
XX genes, their use in the discovery of novel antibiotics, the essential
XX genes themselves and the encoded proteins. The prokaryotes used are
XX Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
XX pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
XX invention is also useful for the identification of potential new targets
XX to identify proteins used in proliferation, to express these proteins,
XX and to obtain antibodies capable of binding to the expressed proteins.
XX The proteins can be used to screen compounds in rational drug discovery
XX programmes. The antisense nucleic acid sequence is also useful to screen
XX for homologous nucleic acids which are required for cell proliferation in
XX essential prokaryotic cellular proliferation protein.
XX Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic
XX format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 202 AA;

Query Match 48.5%; Score 465.5; DB 22; Length 202;
Best Local Similarity 46.5%; Pred. No. 3.8e-44;
Matches 94; Conservative 40; Mismatches 59; Indels 9; Gaps 2;

Qy 1 MLIALIILAYLIGSIPSLVGLKAGIDIREHSGNLGATNAPRTLGKAGSVVIAGD 60
Db 1 MMIIIVMLLSYLIGAFPSGFVIGLFFKKDIRQFGSGNTGATNSFRVLGRPAGFLVTFD 60

Qy 61 ILKGTALATAPFLM--HVD-----THPLLAGVFAVLGHVFPFAKFGKAVATSGGV 111
Db 61 IFKGFITVFFLWLPVHADGPISFTFTNGLIVGLFAILGHVYVPLAFQGGKAVATSGV 120

Qy 112 LLFYAPLLFITWAVFETFLVLTFRVSLSSMLTGIYTVIYFFVHDYLLIVVTLITFV 171
Db 121 VLGNPILLLLALIFFVLKIFVSLASIAAICCVIGSLIITODYILLVSVLSLIL 180

Qy 172 IYHRANIKRIINTEPKVKWL 193
Db 181 IIRHSNISIRFGEPEKIKWM 202

RESULT 11
ABP26818
ID ABP26818 standard; Protein; 212 AA.
XX AC ABP26818;
XX DT 02-JUL-2002 (first entry)
XX DE Streptococcus polypeptide SEQ ID NO 2812.
XX KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
XX KW group A streptococcus; Streptococcus pyogenes; antibacterial;
XX KW antiinflammatory; infection; vaccine; meningitis; gene therapy.
XX OS Streptococcus agalactiae.
XX PN WO200234771-A2.
XX PD 02-MAY-2002.
XX PF 29-OCT-2001; 2001WO-GB04789.
XX PR 27-OCT-2000; 2000GB-0026333.
XX PR 24-NOV-2000; 2000GB-0028727.
XX PR 07-MAR-2001; 2001GB-0005640.
XX PA (CHIR-) CHIRON SPA.
XX PA (GENO-) INST GENOMIC RES.
XX PI Telford J, Massignani V, Margarit Ros YI, Grandi G, Fraser C;
XX PI Tettelin H;
XX DR WPI; 2002-352536/38.
XX DR N-PSDB; ABN67449.
XX CC New Streptococcus protein for the treatment or prevention of infection
XX CC for detecting a compound that binds to the protein -
XX PS Claim 1; Page 3432; 4525pp; English.
XX CC The invention relates to a protein (ABP25413-ABP30895) from group B
XX CC streptococcus/GAS (Streptococcus agalactiae) or group A streptococcus/GAS
XX CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
XX CC the specification. The proteins have antibacterial and antiinflammatory
XX CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
XX CC antibodies that bind (I) are used in the manufacture of medicaments for
XX CC the treatment or prevention of infection or disease caused by
XX CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
XX CC Nucleic acids encoding (I) are used to detect Streptococcus in a
XX CC biological sample. (I) is used to determine whether a compound binds to

CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC used as a vaccine or diagnostic composition. The disease caused by
CC Streptococcus that is prevented or treated may be meningitis. Nucleic
CC acid encoding (I) may be used to recombinantly produce (I) and may be
CC used in gene therapy. Antibodies to (I) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
CC Streptococcus proteins.
XX Sequence 212 AA;
SQ

Query Match 47.3%; Score 454; DB 23; Length 212;
Best Local Similarity 48.2%; Pred. No. 7.9e-43;
Matches 95; Conservative 36; Mismatches 56; Indels 10; Gaps 3;

Qy 1 MLIALIILAYLIGSIPSLVGLKAGIDIREHSGNLGATNAPRTLGKAGSVVIAGD 60
Db 1 MNIIIMIIIIAYLLGSIQTGLWIGKYFYQVNLRHQSGNTGTNTFRILGKAGIVTITD 60

Qy 61 ILKGTALATAPFLMHV-DIHPLLAGVFAVLGHVFPFAKFGKAVATSGVLLFYAPLL 119
Db 61 ILKGTALATAPFLMHV-DIHPLLAGVFAVLGHVFPFAKFGKAVATSGVLLFYAPLL 120

Qy 120 FITWAVFFIFLYLTFRVSLSSM---LTGIYTVIYFFV-----HDTYLLIVVTLITF 170
Db 121 FLVLLVIFLLTLYLFSMISLSSITVAVVGLSVLIFPLVGLFILLTDYDMIFTTVVILMALT 180

Qy 171 VIYHRANIKRIINTE 187
Db 181 IIRHODNIKRIKROE 197

RESULT 12
ABP26819
ID ABP26819 standard; Protein; 240 AA.
XX AC ABP26819;
XX DT 02-JUL-2002 (first entry)
XX DE Streptococcus polypeptide SEQ ID NO 2814.
XX KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
XX KW group A streptococcus; Streptococcus pyogenes; antibacterial;
XX KW antiinflammatory; infection; vaccine; meningitis; gene therapy.
XX OS Streptococcus pyogenes.
XX PN WO200234771-A2.
XX PD 02-MAY-2002.
XX PF 29-OCT-2001; 2001WO-GB04789.
XX PR 27-OCT-2000; 2000GB-0026333.
XX PR 24-NOV-2000; 2000GB-0028727.
XX PR 07-MAR-2001; 2001GB-0005640.
XX PA (CHIR-) CHIRON SPA.
XX PA (GENO-) INST GENOMIC RES.
XX PI Telford J, Massignani V, Margarit Ros YI, Grandi G, Fraser C;
XX PI Tettelin H;
XX DR WPI; 2002-352536/38.
XX DR N-PSDB; ABN67450.
XX CC New Streptococcus protein for the treatment or prevention of infection
XX CC or disease caused by Streptococcus bacteria, such as meningitis, and
XX CC for detecting a compound that binds to the protein -
XX PS Claim 1; Page 3432; 4525pp; English.
XX CC The invention relates to a protein (ABP25413-ABP30895) from group B

CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS.
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in
 CC the specification. The proteins have antibacterial and antiinflammatory
 CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
 CC antibodies that bind (I) are used in the manufacture of medicaments for
 CC the treatment or prevention of infection or disease caused by
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a
 CC biological sample. (I) is used to determine whether a compound binds to
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
 CC used as a vaccine or diagnostic composition. The disease caused by
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic
 CC acid encoding (I) may be used to recombinantly produce (I) and may be
 CC used in gene therapy. Antibodies to (I) are used for affinity
 CC chromatography, immunoassays, and distinguishing/identifying
 CC Streptococcus proteins.
 XX
 SQ Sequence 240 AA;
 Query Match 47.28; Score 453; DB 23; Length 240;
 Best Local Similarity 47.08; Pred. No. 1.2e-42;
 Matches 95; Conservative 34; Mismatches 63; Indels 10; Gaps 3;
 QY 1 MLIALLIAYLIGSIPGLIVGKLAKGIDIREHSGNGLGATNAPRTTGVKAGSVVIAGD 60
 DB 28 MKLLFTYIALIGSIPGLIGQVYFHINREHSGNGTGTNTFRILGVKAGTATLAD 87
 QY 61 ILKGTALVALPFLMHV-DIHPLLAGVFAVLGHVFPFAKFKGKAVATSGGVLLFYAPLL 119
 DB 88 MFKGTLSLLPIIFOMTSISSIAIGFFAVLGHVTPFIFANFKGKAVATSGVLLGFAPLY 147
 QY 120 FITWAVAFIFLYLTKFVLSLM---LTGIYTVIYSFFVH-----DYLLIVVTLTIF 170
 DB 148 LFFLASIEVLVLYFSMISLASVSAIVGSLVLTFFPAIHLLPNYDYELTFIVILLAFI 207
 QY 171 VYHRANIKRIINKTEPKVKW 192
 DB 208 IIRHKONISRIKHTENLIPW 229
 RESULT 13
 ABP30349
 ID ABP30349 standard; Protein; 207 AA.
 XX
 AC ABP30349;
 XX
 DT 02-JUL-2002 (first entry)
 XX
 DE Streptococcus polypeptide SEQ ID NO 9874.
 XX
 KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
 KW group A streptococcus; Streptococcus pyogenes; antibacterial;
 KW antiinflammatory; infection; vaccine; meningitis; gene therapy.
 OS Streptococcus agalactiae.
 XX
 PN WO200234771-A2.
 XX
 PD 02-MAY-2002.
 XX
 PF 29-OCT-2001; 2001WO-GB04789.
 XX
 PR 27-OCT-2000; 2000GB-0026333.
 PR 24-NOV-2000; 2000GB-0028727.
 PR 07-MAR-2001; 2001GB-0005640.
 XX
 PA (CHIR-) CHIRON SPA.
 PA (GENO-) INST GENOMIC RES.
 XX
 PI Telford J, Massignani V, Margarit Ros Yi, Grandi G, Fraser C;
 PI Tettelin H;
 PI
 XX WPI; 2002-352536/38.

DR N-PSDB; ABN70980.
 XX
 PT New Streptococcus protein for the treatment or prevention of infection
 PT or disease caused by Streptococcus bacteria, such as meningitis, and
 PT for detecting a compound that binds to the protein.
 XX
 PS Claim 1; Page 4116; 4525pp; English.
 XX
 CC The invention relates to a protein (ABP25413-ABP30895) from group B
 CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in
 CC the specification. The proteins have antibacterial and antiinflammatory
 CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
 CC antibodies that bind (I) are used in the manufacture of medicaments for
 CC the treatment or prevention of infection or disease caused by
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a
 CC biological sample. (I) is used to determine whether a compound binds to
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
 CC used as a vaccine or diagnostic composition. The disease caused by
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic
 CC acid encoding (I) may be used to recombinantly produce (I) and may be
 CC used in gene therapy. Antibodies to (I) are used for affinity
 CC chromatography, immunoassays, and distinguishing/identifying
 CC Streptococcus proteins.
 XX
 SQ Sequence 207 AA;
 Query Match 46.68; Score 447; DB 23; Length 207;
 Best Local Similarity 48.48; Pred. No. 4.7e-42;
 Matches 93; Conservative 35; Mismatches 54; Indels 10; Gaps 3;
 QY 6 LILAYLIGSIPGLIVGKLAKGIDIREHSGNGLGATNAPRTTGVKAGSVVIAGDILKGT 65
 DB 1 MILLAYLIGSIPGLIGWIKYQVNLQHGSGNGTNTTFRILGVKAGIVTLTIDILKGT 60
 QY 66 LATALPFLMHV-DIHPLLAGVFAVLGHVFPFAKFKGKAVATSGGVLLFYAPLLFTMV 124
 DB 61 LATLIPIILGITTVSPFFIGFAIIGHTFPFAQFKGKAVATSGAGVLLGAPSFYLL 120
 QY 125 AVFFELYLTKFVLSLM---LTGIYTVIYSFFV-----HDTYLLIVVTLTIFVIYRH 175
 DB 121 VIFLLTYLFSMISLSITVAVVGILSVLIPPLVGLFILTIDYDWTFTTVILMALTIIIRH 180
 QY 176 RANIKRIINKTE 187
 DB 181 QDNKIRKROE 192
 RESULT 14
 ABB54294
 ID ABB54294 standard; Protein; 213 AA.
 XX
 AC ABB54294;
 XX
 DT 16-MAY-2002 (first entry)
 XX
 DE Lactococcus lactis protein ykac.
 XX
 KW Lactococcus lactis protein ykac.
 XX
 OS Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.
 XX
 PN Lactococcus lactis IL1403.
 PN FR2807446-A1.
 XX
 PD 12-OCT-2001.
 XX
 PF 11-APR-2000; 2000FR-0004630.
 XX
 PR 11-APR-2000; 2000FR-0004630.
 XX
 PA (INRG) INRA INST NAT RECH AGRONOMIQUE.
 XX

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run On: December 26, 2002, 00:31:11 ; Search time 300.909 Seconds
(without alignments)
61.660 Million cell updates/sec

Title: US-10-068-080-3
Perfect score: 960
Sequence: 1 MLIALIILAVLIGSIPSGL.....RHRANKRIINTEPKVKWL 193

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	960	100.0	193	1	A69892	conserved hypothet
2	568.5	59.2	198	2	A81598	conserved hypothet
3	566.5	59.0	198	2	AD1235	B. subtilis ynes p
4	468	48.8	213	2	C95098	conserved hypothet
5	468	48.8	213	2	C97966	conserved hypothet
6	464.5	48.4	202	2	A89911	conserved hypothet
7	445	46.4	213	2	B86747	conserved hypothet
8	352.5	36.7	222	1	S75137	hypothetical prote
9	339.5	35.4	226	2	AC1868	hypothetical prote
10	333	34.7	202	2	A82737	conserved hypothet
11	333	34.7	205	2	H97517	hypothetical prote
12	322	33.5	206	2	G83854	hypothetical prote
13	318.5	33.2	239	2	S73812	hypothetical prote
14	309.5	32.2	196	2	C72253	conserved hypothet
15	302	31.5	192	2	E70359	conserved hypothet
16	301	31.4	201	2	AC3593	probable integral
17	299	31.1	208	2	B82313	conserved hypothet
18	296	30.8	198	2	A75324	conserved hypothet
19	286	29.8	216	2	AB0080	probable membrane
20	284.5	29.6	218	2	E87552	conserved hypothet
21	284.5	29.6	239	1	C64227	hypothetical prote
22	271.5	28.3	198	2	C37208	uncharacterized co
23	270.5	28.2	257	2	A82894	conserved hypothet
24	263.5	27.4	202	2	D81378	probable integral
25	261.5	27.2	224	2	G90548	conserved hypothet
26	257.5	26.8	203	2	AF0892	probable membrane
27	257	26.8	200	2	C81894	probable integral
28	257	26.8	200	2	H81126	conserved hypothet
29	255	26.6	199	2	E64146	hypothetical prote

30	252	26.2	189	2	A83573	conserved hypothet:
31	250.5	26.1	205	1	A65094	ylGH protein - Esc
32	250.5	26.1	205	2	F91121	hypothetical prote
33	250.5	26.1	205	2	E85966	hypothetical prote
34	230	24.0	220	2	B71811	hypothetical prote
35	228	23.8	262	1	E64708	conserved hypothet
36	196.5	20.7	203	2	F72273	conserved hypothet
37	165	17.2	195	2	A75295	conserved hypothet
38	141	14.7	559	2	B75477	conserved hypothet
39	115	12.0	463	2	A69905	conserved hypothet
40	110.5	11.5	421	2	G72300	conserved hypothet
41	104.5	10.9	963	2	AF2119	hypothetical prote
42	100	10.4	374	2	F90361	transport membrane
43	98.5	10.3	276	2	A75037	hypothetical prote
44	97.5	10.2	322	2	F95970	probable sugar upt
45	97.5	10.2	351	1	C69198	phospho-N-acetylmu

ALIGNMENTS

RESULT 1
A69892
conserved hypothetical protein ynes - Bacillus subtilis
C:Species: Bacillus subtilis
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C:Accession: A69892
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.;
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrar
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.;
Iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lard
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.;
Y., M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Port
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, M.; Rose, M.; Sadale, Y.; Sato, T.; Sc
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.;
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terptrata, P.; Togononi, A.; Tosato, V.; Uch
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yosh
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subt
A:Reference number: A69580; MUID:98044033; PMID:9384377
A:Accession: A69892
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-193 <KUN>
A:Cross-references: GB:Z99113; GB:AL009126; NID:g2634090; PIDN:CAB13690.1; PID:g26
A:Experimental source: strain 168
C:Genetics:
A:Gene: ynes
C:Superfamily: Escherichia coli ygiH protein

Query Match 100.0%; Score 960; DB 1; Length 193;
Best Local Similarity 100.0%; Pred. No. 5.2e-69;
Matches 193; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MLIALIILAVLIGSIPSGLVGKLAKGIDIREHSGNGLGATNAFTLGVRAGSVYIAGD	60
Db	1	MLIALIILAVLIGSIPSGLVGKLAKGIDIREHSGNGLGATNAFTLGVRAGSVYIAGD	60
Qy	61	ILKGTALATLPLMHVDHPLLAGVFAVLGHVFPFAFKGKAVATSGVLLFYAPLLF	120
Db	61	ILKGTALATLPLMHVDHPLLAGVFAVLGHVFPFAFKGKAVATSGVLLFYAPLLF	120
Qy	121	ITWVAFFIFLYLTKFVSLSSMLTGIYTVIYVSFFVHDYVLLIVTLLTIFVYRHRANIK	180
Db	121	ITWVAFFIFLYLTKFVSLSSMLTGIYTVIYVSFFVHDYVLLIVTLLTIFVYRHRANIK	180
Qy	181	RIINTEPKVKWL	193
Db	181	RIINTEPKVKWL	193

RESULT 2

AB1598
C:Species: conserved hypothetical protein, B. subtilis ynes protein homolog lin1323 [imported] - L
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
Science 294, 849-852, 2001
A:Authors: Krefit, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A:Title: Comparative genomics of *Listeria* species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AB1598
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-198 <GLA>
A:Cross-references: GB:AL592022; PIDN:CAC96554.1; PID:g16413796; GSPDB:GN00178
A:Experimental source: strain Clilp1262
C:Genetics:
A:Gene: lin1323
C:Superfamily: Escherichia coli ygiH protein

Query Match 59.2%; Score 568.5; DB 2; Length 198;
Best Local Similarity 57.7%; Pred. No. 4.7e-38;
Matches 112; Conservative 35; Mismatches 44; Indels 3; Gaps 1;
Qy 3 IALLIILAYLIGSIPSLVGLKAGIDIREHSGNLGATNFRTLGVKAGSVIAGDIL 62
Db 5 LILLSLLAYVIGSIPSLVGLKAGIDIREHSGNLGATNFRTLGVKAGSVIAGDIL 64
Qy 63 KGTALATLPEL--MHVDIH-PLLAGVFAVLGHVPIFAKFKGKAVATSGGVLLFYAPLL 119
Db 65 KGTATLALPFLMHVD--IHPLLAGVFAVLGHVPIFAKFKGKAVATSGGVLLFYAPLL 124
Qy 120 FITWAVFFILYLTKFVLSLMLTGIVTYVVSFFVHDYLLIVVTLTLTIFVIYHRANI 179
Db 125 FVAALVFLVLTKLKSYVLSLMLTGIVTYVVSFFVHDYLLIVVTLTLTIFVIYHRANI 184
Qy 180 KRIINKTEPKVKWL 193
Db 185 TRIRNGEPEKIKWM 198

RESULT 3

AD1235
B. subtilis ynes protein homolog lml284 [imported] - *Listeria* monocytogenes (strain EGD
C:Species: *Listeria* monocytogenes
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
Science 294, 849-852, 2001
A:Authors: Krefit, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A:Title: Comparative genomics of *Listeria* species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AD1235
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-198 <GLA>
A:Cross-references: GB:NC_003210; PIDN:CAC99362.1; PID:g16410700; GSPDB:GN00177
A:Experimental source: strain EGD-e
C:Genetics:
A:Gene: lml284
C:Superfamily: Escherichia coli ygiH protein

Query Match 59.0%; Score 566.5; DB 2; Length 198;
Best Local Similarity 58.8%; Pred. No. 6.8e-38;
Matches 114; Conservative 33; Mismatches 44; Indels 3; Gaps 2;

Qy 3 IALLIILAYLIGSIPSLVGLKAGIDIREHSGNLGATNFRTLGVKAGSVIAGDIL 62
Db 5 LILLSLLAYVIGSIPSLVGLKAGIDIREHSGNLGATNFRTLGVKAGSVIAGDIL 64
Qy 63 KGTALATLPEL--MHVDIH-PLLAGVFAVLGHVPIFAKFKGKAVATSGGVLLFYAPLL 119
Db 65 KGTATLALPFLMHVD--IHPLLAGVFAVLGHVPIFAKFKGKAVATSGGVLLFYAPLL 124
Qy 120 FITWAVFFILYLTKFVLSLMLTGIVTYVVSFFVHDYLLIVVTLTLTIFVIYHRANI 179
Db 125 FVAALVFLVLTKLKSYVLSLMLTGIVTYVVSFFVHDYLLIVVTLTLTIFVIYHRANI 184
Qy 180 KRIINKTEPKVKWL 193
Db 185 TRIRNGEPEKIKWM 198

RESULT 4

C95098
conserved hypothetical protein SP0851 [imported] - *Streptococcus pneumoniae* (strain
C:Species: *Streptococcus pneumoniae*
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
C:Accession: C95098
R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.;
nson, T.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzaj
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Mori
A:Title: Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: C95098
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-213 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK74980.1; PID:g14972323; GSPDB:GN00164; TI
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SP0851
C:Superfamily: Escherichia coli ygiH protein

Query Match 48.8%; Score 468; DB 2; Length 213;
Best Local Similarity 48.0%; Pred. No. 4.4e-30;
Matches 97; Conservative 32; Mismatches 63; Indels 10; Gaps 3;

Qy 1 MLIALIILAYLIGSIPSLVGLKAGIDIREHSGNLGATNFRTLGVKAGSVIAGD 60
Db 1 MITIVLLILAYLIGSIPSLVGLKAGIDIREHSGNLGATNFRTLGVKAGSVIAGD 60
Qy 61 ILKGTALATLPELHMVD--IHPLLAGVFAVLGHVPIFAKFKGKAVATSGGVLLFYAPLL 119
Db 61 FFKGTALATLPIIFHLOGVSPILFGLLAVIGHTPTFPAKFKGKAVATSGGVIFGAPIF 120
Qy 120 FITWAVFFILYLTKFVLSLMLTGIVTYVVSFFVHDYLLIVVTLTLTIF 170
Db 121 CLYLAIIFGALYLGSMISLSVTSASIAAVIGVLLFPLFGFILSNYDSLTAILALASL 180
Qy 171 VIYHRANIKRIINKTEPKVKW 192
Db 181 IIRHKDNRIARKNTENLVPM 202

RESULT 5

C97966
conserved hypothetical protein spr0755 [imported] - *Streptococcus pneumoniae* (stra
C:Species: *Streptococcus pneumoniae*
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
C:Accession: C97966
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren,
J. P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas,

Db 66 ISKGVMAVALRAIYSGDMLPALPAAMONWLTGVAIAVVLGHSHKSFIFKFSCKGKSVATS 125
Qy 109 GGVLLFYAPLLFITWVAFFFLYLVKFSVSSMLTGTYT--VIYSFFVHDYLLIVVTL 166
Db 126 LGLVFLMLNLALGTATLFLVIFTRIVSSSIVAALAVNGIALALQPPYLAFTF-L 184
Qy 167 LTFIVYIYHRANIKRIINKTEPKV 190
Db 185 AGMYVIVRHRNTRNIHQTEPKL 208

RESULT 9

AC1868
hypothetical protein all0492 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp.
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: AC1868
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriouchi, N.; Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AC1868
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-226 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA072450.1; PID:gl7129837; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Gene: all0492
C:Superfamily: Escherichia coli ygiH protein

Query Match 35.4%; Score 339.5; DB 2; Length 236;
Best Local Similarity 40.2%; Pred. No. 6.7e-20;
Matches 84; Conservative 37; Mismatches 87; Indels 21; Gaps 5;
Qy 2 LIALLIILAYLIGSPGLVGLKAGIDIREHSGNGLGATNAPRTLVGKAGSVVIAGDI 61
Db 7 LCGAVLVAYLLGSPPTGYIAVKQLKIDIREVSGSGTATNVLRTLGKGPFAVLGLDC 66
Qy 62 LKGTATLAL-----PFLMHVDI-HP---LLAGVFAVLGHVFFIFAKFKGK 103
Db 67 LKGVLAIALVYLFASFASQNLPIPTVNIELQWPLVTLGAILGHSKIFLGTGK 126
Qy 104 AVATSGGVLLFYAPLLFITWVAFFFLYLVKFSVSSMLTGTYT--VIYSFFVHD--TYLL 161
Db 127 SVATSLGILLANWQVGLATGCVFAVVVAISRVLSLSSIMGAIVSVWFLOQLPYIL 186
Qy 162 IVVTLTIFVIYHRANIKRIINKTEPKV 190
Db 187 FGIA-GGLYVILRHSNRIERLAGTEPKI 214

RESULT 10

AB2737
conserved hypothetical protein Atul306 [imported] - Agrobacterium tumefaciens (strain C58)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 01-Feb-2002
C:Accession: AB2737
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.; Erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClellan, K.; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, S.E.; E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AB2737
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-202 <KUR>

A:Cross-references: GB:AE008688; PIDN:AA42312.1; PID:gl7739714; GSPDB:GN00186
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atul306
A:Map position: circular chromosome
C:Superfamily: Escherichia coli ygiH protein

Query Match 34.7%; Score 333; DB 2; Length 202;
Best Local Similarity 38.1%; Pred. No. 2e-19;
Matches 72; Conservative 43; Mismatches 74; Indels 0; Gaps 0;

Qy 2 LIALLIILAYLIGSPGLVGLKAGIDIREHSGNGLGATNAPRTLVGKAGSVVIAGDI 61
Db 10 LLAALALIGVLLGSPFGLILTRMAGLDVRKIGSGNIGATNVLRTGNKLAATLLDA 69
Qy 62 LKGTATLALPFLMHVDIHPHLLAGVFAVLGHVFFIFAKFKGKAVATSGGVLLFYAPLLFI 121
Db 70 LKGTAAVLVANALWGYESLVAGFFAFGLHLPVWLFGKGVAVVIGVLLGNAAPLML 129
Qy 122 TMVAFFIFLYLVKFSVSSMLTGTYT--VIYSFFVHDYLLIVVTLTIFVIYHRANIKR 181
Db 130 AFALILWLTATFITRYSSLSALLAMLIIPVALWLGPEKTAMLVTLSSVISWKKHREIRR 189
Qy 182 IINKTEPKV 190
Db 190 LMACTESRI 198

RESULT 11

H97517
hypothetical protein AGR_C_2402 [imported] - Agrobacterium tumefaciens (strain C58).
C:Species: Agrobacterium tumefaciens
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
C:Accession: H97517
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Go:
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacteri
A:Reference number: A97359; PMID:11743194
A:Accession: H97517
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-205 <KUR>
A:Cross-references: GB:AE007869; PIDN:AAK87097.1; PID:gl5156359; GSPDB:GN00169
C:Genetics:
A:Gene: AGR_C_2402
A:Map position: circular chromosome
C:Superfamily: Escherichia coli ygiH protein

Query Match 34.7%; Score 333; DB 2; Length 205;
Best Local Similarity 38.1%; Pred. No. 2e-19;
Matches 72; Conservative 43; Mismatches 74; Indels 0; Gaps 0;

Qy 2 LIALLIILAYLIGSPGLVGLKAGIDIREHSGNGLGATNAPRTLVGKAGSVVIAGDI 61
Db 13 LLAALALIGVLLGSPFGLILTRMAGLDVRKIGSGNIGATNVLRTGNKLAATLLDA 72
Qy 62 LKGTATLALPFLMHVDIHPHLLAGVFAVLGHVFFIFAKFKGKAVATSGGVLLFYAPLLFI 121
Db 73 LKGTAAVLVANALWGYESLVAGFFAFGLHLPVWLFGKGVAVVIGVLLGNAAPLML 132
Qy 122 TMVAFFIFLYLVKFSVSSMLTGTYT--VIYSFFVHDYLLIVVTLTIFVIYHRANIKR 181
Db 133 AFALILWLTATFITRYSSLSALLAMLIIPVALWLGPEKTAMLVTLSSVISWKKHREIRR 192
Qy 182 IINKTEPKV 190
Db 193 LMACTESRI 201

RESULT 12

G83854
hypothetical protein BH1639 [imported] - Bacillus halodurans (strain C-125)

C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: G83854
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20512582; PMID:11058132
C:Accession: G83854
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-206 <STO>
A:Cross-references: GB:AF001512; GB:BA000004; NID:g10174030; PIDN:BA05358.1; GSPDB:GN00
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH1639
C:Superfamily: Escherichia coli ygiH protein

Query Match 33.5%; Score 322; DB 2; Length 206;
Best Local Similarity 37.3%; Pred. No. 1.5e-18;
Matches 75; Conservative 43; Mismatches 61; Indels 22; Gaps 4;

Qy 5 LLIILYLGISPSGLIVGKLAGIDIREHSGNLTATNAPFTLGVKAGSVVIAGDILKG 64
Db 6 LLIIVSYLGISVFSYIIIAIKKIDIRQHGSGNAGATNLRVLGVPATVLLDILKG 65

Qy 65 TLATLALPFLMHVDIHP-----LLAGVFAVLGHVFPFIFAKFGKCAVATSGVLLFYAP 117
Db 66 VTAV-----VTVLTDPDGGWFAAAGIAIIGHNWPYIGFRGKGATTVIGVLAASLP 121

Qy 118 LLEFI-----TWAVFFIFLYLKFSLSMLTGIVTVIYSFVVDHTYLLIVTLTI 169
Db 122 LAALVAGTAVIGSIWTVRVSLGSLFTVLTALLAVLSQWFGVPVYIITLIIIVAIL- 180

Qy 170 FVIYHRANIKRIINKTEPKV 190
Db 181 --MWRHRSIQRLSLGCTENKL 199

RESULT 13
S73812
hypothetical protein ygiH - Mycoplasma pneumoniae (strain ATCC 29342)
N:Alternate names: hypothetical protein H91_orf239
C:Species: Mycoplasma pneumoniae
A:Variety: ATCC 29342
C:Date: 27-Feb-1997 #sequence_revision 25-Apr-1997 #text_change 07-Dec-1999
C:Accession: S73812
R:Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkel, E.; Li, B.C.; Herrmann, R.
Nucleic Acids Res. 24, 4420-4449, 1996
A:Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae
A:Reference number: S73327; MUID:97105885; PMID:8948633
A:Accession: S73812
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-239 <HIM>
A:Cross-references: EMBL:AE000047; GB:U00089; NID:g1674162; PIDN:AAB96134.1; PID:g167417
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996
C:Genetics:
A:Gene: ygiH
A:Genetic code: SGC3
C:Superfamily: Escherichia coli ygiH protein

Query Match 33.2%; Score 318.5; DB 2; Length 239;
Best Local Similarity 34.2%; Pred. No. 3.2e-18;
Matches 79; Conservative 45; Mismatches 68; Indels 39; Gaps 6;

Qy 1 MLTALLIILAYLIGSPGLIVGKLAGIDIREHSGNLTATNAPFTLGVKAGSVVIAGD 60
Db 9 LLIVSLVIGYLMGSMFADFVKGILNK-DVRKLGSKNPGATNSIRVGLKIGFLVGLCD 67

Qy 61 ILKGTALATPLFM-----HVDIH-----PLLAGVFAVLGHVFPFIFAKFGKAV 105
Db 68 ALKGLFAVFSFLTSFVLMQYLLNVNQYKVVYLYLSQFAATIGHIFPLFKPKGKRAI 127

Qy 106 ATSGVLLFYAPLFFITMVAVFIFLYLTKEVSLSSMLTGIIYTVI-----YSFFV-- 155
Db 128 ATTGCSLLAISLWFWFICLVLLWLLVTLITKYVSLASLWFFIFLAILIILVPHLDLYFKP 187

Qy 156 -----HDTYLLIVVTL-----LTIFVIYHRANIKRIINKTEPKVKWL 193
Db 188 NPINAISYQNDWYIILFVFLWYPLTIAVFWLHRKNIHRLLENKTKENKVTL 238

RESULT 14
C72253
conserved hypothetical protein - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: C72253
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.;
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardso
C.M.
Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from geno
A:Reference number: A72200; MUID:99287316; PMID:10360571
A:Accession: C72253
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-196 <ARN>
A:Cross-references: GB:AE001796; GB:AE000512; NID:g4982004; PIDN:AAD36515.1; PID:g
A:Experimental source: strain MSB8
C:Genetics:
A:Gene: TM1447
C:Superfamily: Escherichia coli ygiH protein

Query Match 32.2%; Score 309.5; DB 2; Length 196;
Best Local Similarity 37.6%; Pred. No. 1.4e-17;
Matches 71; Conservative 34; Mismatches 79; Indels 5; Gaps 3;

Qy 5 LLIILAYLIGSPGLIVGKLAGIDIREHSGNLTATNAPFTLGVKAGSVVIAGDILKG 64
Db 5 LFPILGYFIGSIPSPSYLIPKWLKIDVRKVGSGNVTATRTTGPVGGICLLDALKG 64

Qy 65 TLATLALPFLMHVDIHP--LAGVFAVLGHVFPFIFAKFGKCAVATSGVLLFYAPLFLITM 123
Db 65 FFPVFTITFGSDKISVLSLTATVGLHDFIFMKFGKGVASTGLIIFCLSNPTGLVF 124

Qy 124 VAVFFIFLYLKFSLSMLTGIVTVIYSFV--HDTYLLIVVTLTIFVIYHRANIKR 181
Db 125 TLTMLVIVMLTKYASLSGLVALYVSALLGYLLKGYDTGMLELI--LAVLSTLRHSENIQR 182

Qy 182 IINKTEPKV 190
Db 183 LLNGTERKV 191

RESULT 15
E70359
conserved hypothetical protein aq_676 - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 20-Sep-1999
C:Accession: E70359
R:Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.;
V.
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:98196666; PMID:9537320
A:Accession: E70359
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-192 <AQF>
A:Cross-references: GB:AE000701; NID:g2983260; PIDN:AAC08869.1; PID:g2983275; GB:A
A:Experimental source: strain VF5
C:Genetics:
A:Gene: aq_676
C:Superfamily: Escherichia coli ygiH protein

```

Query Match      31.58; Score 302; DB 2; Length 192;
Best Local Similarity 37.7%; Pred. No. 5.4e-17;
Matches 75; Conservative 36; Mismatches 66; Indels 22; Gaps 6;

Qy 1 MJIALIILAYLIGSIPSLIVGKLAGIDIREHSGNLGATNAFTLGKAGSWIAGD 60
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 MKALFLVIFAYLLGSITFGEVIAKL-KGVDLRWVSGNVCATNVTALGKKYGVLFVFLD 59
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 61 ILKCTLATALPFLMHVDIHP-----LAGVFAVLGHVFFIFAKFKGKAVATSGGVLLF 114
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 60 FLKG----FIPALIAVKSFSGIDSWLFTFTGLASVLGHMYPVFFGFGKGVATAGLVVFA 115
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 115 YAPLLFITWAVFFIFLYL-----TKFVLSLSSMLTGIYTVIYSFFV-HDTYLLIVVTLT 168
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 116 VSP-----SVALFSFLVWLGIPLWKRYVSLASITATISAFLEFVAGYPVNVLFMAIVIG 170
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 169 IFVIYRHRANIKRIINKTE 187
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 171 ALIYRHRNINRLTGRE 189
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Search completed: December 26, 2002, 01:27:06
Job time : 301.909 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 25, 2002, 14:27:20 ; Search time 137.382 Seconds
(without alignments)
58.268 Million cell updates/sec

Title: US-10-068-080-3
Perfect score: 960
Sequence: 1 MLIALIILAYLIGSIPSGL.....RHRANIKRIINKTEPKVKWL 193

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	960	100.0	193	1	YNES_BACSU	Q45064	bacillus su
2	468	48.8	213	1	Y851_STRPN	Q54916	streptococc
3	352.5	36.7	222	1	Y373_SYNY3	P73933	synchocyst
4	318.5	33.2	239	1	Y247_MYCPN	P73428	mycoplasma
5	302	31.5	192	1	Y576_AQUAE	O66905	aquifex aeo
6	284.5	29.6	239	1	Y247_MYCCE	P47489	mycoplasma
7	255	26.6	199	1	YG1H_HAEIN	P44603	haemophilus
8	250.5	26.1	205	1	YGIH_ECOLI	P31056	escherichia
9	230	24.0	220	1	YF09_HELPJ	Q92jbl	helicobacte
10	228	23.8	220	1	YF09_HELPY	Q26039	helicobacte
11	97.5	10.2	351	1	MRAY_METH	Q26830	methanobact
12	97	10.1	514	1	COX1_PONPA	P26892	pongo pygma
13	92	9.6	514	1	COX1_PIG	O79876	sus scrofa
14	92	9.6	1060	1	NKCL_MANSE	Q25479	manduca sex
15	91.5	9.5	815	1	NAHL_HUMAN	P19634	homo sapien
16	90.5	9.4	380	1	Y420_METJA	Q57863	methanococc
17	90	9.4	513	1	COX1_HUMAN	P00395	homo sapien
18	89.5	9.3	817	1	NAHL_BOVIN	Q28036	bos taurus
19	89	9.3	370	1	EXOH_RHIME	P33692	rhizobium m
20	89	9.3	383	1	NAPA_ENTHR	P26235	enterococcu
21	88	9.2	385	1	Y421_METTH	Q26521	methanobact
22	88	9.2	426	1	BRNQ_CORGL	O06754	corynebacte
23	87.5	9.1	329	1	SRAT_CABEL	Q92029	caenorhabdi
24	87	9.1	516	1	COX1_BALPH	P24983	balaeonopter
25	87	9.1	919	1	PNAL_SCHPO	P09627	schizosacch
26	86.5	9.0	1021	1	TSCC_HUMAN	P50127	homo sapien
27	86	9.0	513	1	COX1_DASNO	O21327	dasyopus nov
28	86	9.0	514	1	COX1_SHEEP	O78749	ovis aries
29	86	9.0	516	1	COX1_SHEEP	P41293	balaeonopter
30	85.5	8.9	274	1	COX1_CHORO	P30671	choristoneu
31	85	8.9	514	1	COX1_BOVIN	P00396	bos taurus
32	84.5	8.8	292	1	MNTC_BACHD	Q9Kd29	bacillus ha
33	84.5	8.8	317	1	HTPX_THEAC	Q9hiv2	thermoplasm

34	84.5	8.8	549	1	COX1_LEITA	P14544	leishmania
35	84.5	8.8	816	1	NAHL_RABIT	P23791	oryctolagus
36	84	8.8	513	1	COX1_DIDMA	P41310	didelphis m
37	84	8.8	513	1	COX1_PAPHA	Q92xy2	papio hamad
38	84	8.8	514	1	COX1_EQUAS	P92477	equus asinu
39	84	8.8	514	1	COX1_HORSE	P48659	equus caball
40	83.5	8.7	274	1	COX1_CHOFU	P50669	choristoneu
41	83.5	8.7	820	1	NAHL_MOUSE	O61165	mus musculus
42	83	8.6	502	1	NU2C_MESVI	Q9nuq6	mesostigma
43	83	8.6	529	1	YABD_SCHPO	Q09812	schizosacch
44	83	8.6	1002	1	TSCC_RAT	P5018	rattus norv
45	82.5	8.6	217	1	OPCB_BACSU	O34878	bacillus su

ALIGNMENTS

RESULT 1	YNES_BACSU	STANDARD;	PRT;	193 AA.
ID	YNES_BACSU	STANDARD;	PRT;	193 AA.
AC	Q45064;	1999 (Rel. 38, Created)		
DT	15-JUL-1999	(Rel. 38, Last sequence update)		
DT	15-JUN-2002	(Rel. 41, Last annotation update)		
DE	Hypothetical protein ynes.			
GN	YNES.			
OS	Bacillus subtilis.			
OC	Bacteria: Firmicutes; Bacillales; Bacillaceae; Bacillus.			
OX	NCBI_TaxID=1423;			
RN	SEQUENCE FROM N.A.			
RP	STRAIN=168;			
RC	MEDLINE=97124194; PubMed=8969507;			
RX	Rose M., Entian K.D.;			
RA	"New genes in the 170 degree region of the Bacillus subtilis genome			
RT	encode DNA gyrase subunits, a thioredoxin, a xylanase and an amino			
RT	acid transporter."			
RL	Microbiology 143:3097-3101(1996).			
RN	SEQUENCE FROM N.A.			
RP	STRAIN=168;			
RC	MEDLINE=98044033; PubMed=9384377;			
RX	Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,			
RA	Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,			
RA	Borries R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,			
RA	Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,			
RA	Choi S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A.,			
RA	Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,			
RA	Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,			
RA	Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,			
RA	Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,			
RA	Guiseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,			
RA	Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,			
RA	Joris B., Karamata D., Kasahara Y., Kleier-Blanchard M., Klein C.,			
RA	Kobayashi Y., Koetter P., Koningstein G., Krogh S., Kumano M.,			
RA	Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,			
RA	Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue G.,			
RA	Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,			
RA	Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,			
RA	Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,			
RA	Presecan E., Puig P., Purnelle B., Rapoport G., Rey M., Reynolds S.,			
RA	Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,			
RA	Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,			
RA	Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,			
RA	Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,			
RA	Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognoni A.,			
RA	Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,			
RA	Viari A., Wambut R., Wedler E., Wedler H., Weitzenegger T.,			
RA	Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,			
RA	Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.,			
RT	"The complete genome sequence of the Gram-positive bacterium Bacillus			
RT	subtilis."			
RL	Nature 390:249-256(1997).			

CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -!- SIMILARITY: BELONGS TO THE UPF0078 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: Z73234; CAA97604.1; -
 CC EMBL: Z99113; CAB13690.1; -
 CC Subtilist: BG11826; YncS.
 CC InterPro: IPR003811; DUF205.
 CC Pfam: PF02660; DUF205; 1.
 CC TIGRFAMS: TIGR00023; DUF205; 1.
 CC KW Hypothetical protein; Transmembrane; Complete proteome.
 FT TRANSMEM 2 22 POTENTIAL.
 FT TRANSMEM 54 74 POTENTIAL.
 FT TRANSMEM 76 96 POTENTIAL.
 FT TRANSMEM 111 131 POTENTIAL.
 FT TRANSMEM 153 173 POTENTIAL.
 SQ SEQUENCE 193 AA; 20966 MW; C75803C399B97292 CRC64;
 Query Match 100.08; Score 960; DB 1; Length 193;
 Best Local Similarity 100.08; Pred. No. 1.6e-63;
 Matches 193; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 MLIALLIAYLIGSLVGVKLGKIDIREHSGNGLGNATNFTLGVKAGSVVIAGD 60
 Db 1 MLIALLIAYLIGSLVGVKLGKIDIREHSGNGLGNATNFTLGVKAGSVVIAGD 60
 Oy 61 ILKGTALATLPLMHVDIHPLLAGVAVLGHVFPFPAKFGKAVATSGVLLFYAPLLF 120
 Db 61 ILKGTALATLPLMHVDIHPLLAGVAVLGHVFPFPAKFGKAVATSGVLLFYAPLLF 120
 Oy 121 ITWAVFFFLYLTKEVSLSSMLTGITYTVIYFSFFVHDYLLIIVTLLFIYIYRHRANIK 180
 Db 121 ITWAVFFFLYLTKEVSLSSMLTGITYTVIYFSFFVHDYLLIIVTLLFIYIYRHRANIK 180
 Oy 181 RIINKTEPKVKWL 193
 Db 181 RIINKTEPKVKWL 193
 RESULT 2
 Y851_STRPN
 ID Y851_STRPN STANDARD; PRT; 213 AA.
 AC Q54916;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein SP0851.
 GN SP0851
 OS Streptococcus pneumoniae.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 CC Streptococcus.
 OX NCBI_TaxID=1313;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=7785;
 RX MEDLINE=96312346; PubMed=8763932;
 RA Pan X., Fisher M.;
 RT "Cloning and characterization of the parC and parE genes of
 RT Streptococcus pneumoniae encoding DNA topoisomerase IV: role in
 RT fluoroquinolone resistance".
 RL J. Bacteriol. 178:4060-4069(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TIGR4;
 RX MEDLINE=21357209; PubMed=11463916;
 RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,

RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
 RA Durkin A.S., White M., Kolonay J.F., Nelson W.C., Peterson J.D.,
 RA Unayam L.A., Whitt O., Salzberg S.L., Lewis M.R., Radune D.,
 RA Holtzapple E., Khouri H., Wolf A.M., Ufferback T.R., Hansen C.L.,
 RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
 RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
 RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
 RT "Complete genome sequence of a virulent isolate of Streptococcus
 RT pneumoniae".
 RL Science 293:498-506(2001).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -!- SIMILARITY: BELONGS TO THE UPF0078 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: Z67739; CAA91549.1; -
 CC EMBL: AE007390; AAK74980.1; -
 CC TIGR: SP0851;
 CC InterPro: IPR003811; DUF205.
 CC Pfam: PF02660; DUF205; 1.
 CC TIGRFAMS: TIGR00023; DUF205; 1.
 CC KW Hypothetical protein; Transmembrane; Complete proteome.
 FT TRANSMEM 2 22 POTENTIAL.
 FT TRANSMEM 70 90 POTENTIAL.
 FT TRANSMEM 112 132 POTENTIAL.
 FT TRANSMEM 143 163 POTENTIAL.
 FT TRANSMEM 164 184 POTENTIAL.
 FT CONFLICT 168 168 S->F (IN REF. 1).
 SQ SEQUENCE 213 AA; 22928 MW; 22CB089C17750818 CRC64;
 Query Match 48.88; Score 468; DB 1; Length 213;
 Best Local Similarity 48.08; Pred. No. 1.1e-27;
 Matches 97; Conservative 32; Mismatches 63; Indels 10; Gaps 3;
 Oy 1 MLIALLIAYLIGSLVGVKLGKIDIREHSGNGLGNATNFTLGVKAGSVVIAGD 60
 Db 1 MITVLLIAYLIGSLVGVKLGKIDIREHSGNGLGNATNFTLGVKAGSVVIAGD 60
 Oy 61 ILKGTALATLPLMHVDIHPLLAGVAVLGHVFPFPAKFGKAVATSGVLLFYAPLL 119
 Db 61 FFKGTALATLPLIHFILQGVSPILFGLAVIGTTPFPAKFGKAVATSGVIFGFADIF 120
 Oy 120 FITWAVFFFLYLTKEVSLSSMLTGITYTVI-----YSEFV--HDTYLLIVTLLTIF 170
 Db 121 CLYLAIFFGALYLGSMISLSSVSTASIAAVIGVLLFPLFGFILSNYDSLFTAILALASL 180
 Oy 171 VIYRHRANIKRIINKTEPKVKW 192
 Db 181 IIRHKONIARIKKNKTENLVFW 202
 RESULT 3
 YJ73_SYNY3
 ID YJ73_SYNY3 STANDARD; PRT; 222 AA.
 AC P73933;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein sll1973.
 GN SLL1973.
 OS Synechocystis sp. (strain PCC 6803).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
 OX NCBI_TaxID=1148;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97061201; PubMed=8905231;
 RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: AE000701; AAC06869.1; --
 DR InterPro: IPR003811; DUF205.
 DR Pfam: PF02660; DUF205; 1.
 DR TIGRFAMs: TIGR00023; DUF205; 1.
 KW Hypothetical protein; Transmembrane; Complete proteome.
 FT TRANSMEM 3 23 POTENTIAL.
 FT TRANSMEM 51 71 POTENTIAL.
 FT TRANSMEM 80 100 POTENTIAL.
 FT TRANSMEM 112 132 POTENTIAL.
 FT TRANSMEM 149 169 POTENTIAL.
 SQ SEQUENCE 192 AA; 20940 MW; EAD53C4016D63C00 CRC64;

Query Match 31.5%; Score 302; DB 1; Length 192;
 Best Local Similarity 37.7%; Pred. No. 1.2e-15;
 Matches 75; Conservative 36; Mismatches 66; Indels 22; Gaps 6;

Qy 1 MLIALIILAVLIGSLIPGLVGLKAKGIDIREHSGNIGATNAPRTLGVRAGSVVIAGD 60
 Db 1 MKALFLVIFAYLGLSITGEVIKAL-KGVDLRNVGSGNIGATNAPRTLGVRAGSVVIAGD 59
 Qy 61 ILKGLTALPFLMHVDIHLPL-----LAGVFAVLGHVFPPIFAKFGKAVATSGGVLLF 114
 Db 60 FLKG-----FIPALIAVKSFGSDSWLTFTGLASVLGHMYPVFFGFGKGVATGALGVVFA 115
 Qy 115 VAPLFIPTWAVFFIFLVL-----TKFVLSLSMLTGIVTVISFV-HDTVLLIVVILLT 168
 Db 116 VSP-----SVAFSLFVLWGLFIMKRYVSLASITATISAFLEFLFVAGVPVNVFLMAIVIG 170
 Qy 169 IFVIYHRHRANIKRIINKE 187
 Db 171 ALIIYHRHNRLTGRE 189

RESULT 6
 Y247_MYCGE
 ID Y247_MYCGE STANDARD; PRT; 239 AA.
 AC P47489;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein MG247.
 GN MG247.
 OS Mycoplasma genitalium.
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
 OX NCBI_TaxID=2057;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 33530 / G-37;
 RX MEDLINE=96026346; PubMed=7569993;
 RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A., Kirschner E.F.,
 RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Sandusky M., Fritchman J.L., Weidman J.F., Saudek D.M., Phillips C.A., Merrick J.M.,
 RA Nguyen D.T., Utterback T.R., Saudek D.M., Saudek D.M., Saudek D.M., Saudek D.M., Saudek D.M., Saudek D.M., Saudek D.M., Saudek D.M., Saudek D.M.,
 RA Tomb J.-F., Dougherty B.A., Bitt K.F., Hu P.-C., Lucier T.S.,
 RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
 RT "The minimal gene complement of Mycoplasma genitalium";
 RL Science 270:397-403(1995).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -1- SIMILARITY: BELONGS TO THE UPF0078 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: U39703; AAC71467.1; --
 DR TIGR: MG247; --
 DR InterPro: IPR003811; DUF205.
 DR Pfam: PF02660; DUF205; 1.
 DR TIGRFAMs: TIGR00023; DUF205; 1.
 KW Hypothetical protein; Transmembrane; Complete proteome.
 FT TRANSMEM 7 27 POTENTIAL.
 FT TRANSMEM 61 81 POTENTIAL.
 FT TRANSMEM 99 119 POTENTIAL.
 FT TRANSMEM 135 155 POTENTIAL.
 FT TRANSMEM 159 179 POTENTIAL.
 FT TRANSMEM 199 219 POTENTIAL.
 SQ SEQUENCE 239 AA; 27489 MW; D78CE976DEF621FD CRC64;

Query Match 29.6%; Score 284.5; DB 1; Length 239;
 Best Local Similarity 32.2%; Pred. No. 2.8e-14;
 Matches 75; Conservative 43; Mismatches 72; Indels 43; Gaps 6;

Qy 3 IALLIILA-----YLIGSLIPGLVGLKAKGIDIREHSGNIGATNAPRTLGVRAGSVVIA 58
 Db 7 IAILVIFSLASGYLLGSITFIADIFSKILKK-NVREFGSKNPGATNSMRVFLKIGFLVAI 65
 Qy 59 GDILKGLTALPFLM-----HVDIHLPLAGVFAVLGHVFPPIFAKFGK 103
 Db 66 FDAFKGFFAFLLTWILFRFGLQGYLTKFYSLSLMLTGIVTVI-----YSFFV 155
 Qy 104 AVATSGGVGLFYAPLLFITWAVFFIFLYLTKFYSLSLMLTGIVTVI-----YSFFV 155
 Db 126 AIATGSGSLAISLWFLICLLIWMITLTKYVSLASLITFFVLAVILLIPWLDLYLFF 185
 Qy 156 -----HDYLLIVTVL-----LTIFVIYHRHRANIKRIINKEPKVKWL 193
 Db 186 NSDPLKSLITYQNEWYIILFCLWYVPLTVVFWLHRLHRIILHKGESKITQL 238

RESULT 7
 YGIH_HAEIN
 ID YGIH_HAEIN STANDARD; PRT; 199 AA.
 AC P44603;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein HI0266.
 GN HI0266.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OX Haemophilus.
 OX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RJ / KW20 / ATCC 51907;
 RX MEDLINE=95350630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirschner E.F.,
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McKenney K., Sutton G., Fritchman J.L., Weidman J.F., Saudek D.M., Saudek D.M., Saudek D.M., Saudek D.M., Saudek D.M., Saudek D.M., Saudek D.M., Saudek D.M., Saudek D.M.,
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Saudek D.M., Saudek D.M., Saudek D.M., Saudek D.M., Saudek D.M., Saudek D.M., Saudek D.M., Saudek D.M.,
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
 RD";
 RL Science 269:496-512(1995).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -1- SIMILARITY: BELONGS TO THE UPF0078 FAMILY. STRONG, TO E.COLI YGIH.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

Db 182 ORLRRQETKI-W 193

RESULT 9

YF09_HELPJ STANDARD: PRT; 220 AA.
AC Q92JBI:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein JHP1402.
GN JHP1402.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=85963;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guld B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori";
RL Nature 397:176-180(1999).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE UPF0078 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AE001562; AAC06983.1;
DR InterPro: IPR003811; DUF205.
DR Pfam: PF02660; DUF205; 1.
DR TIGRFAMs: TIGR00023; DUF205; 1.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 11 31 POTENTIAL.
FT TRANSMEM 34 54 POTENTIAL.
FT TRANSMEM 70 90 POTENTIAL.
FT TRANSMEM 128 148 POTENTIAL.
FT TRANSMEM 153 173 POTENTIAL.
SQ SEQUENCE 220 AA; 23804 MW; 7FF48CA2202AC023 CRC64;

Query Match 24.08; Score 230; DB 1; Length 220;

Best Local Similarity 31.68; Pred. No. 2.4e-10;
Matches 65; Conservative 38; Mismatches 83; Indels 20; Gaps 6;

QY 5 LLIIILYLGISGPSGLIVGKLAGIDIREHSGNLTGATNAPRTLGVRAGS-----VVI 57
DB 14 IFTLLGLIGIPGYALMKIFYGMDITKIGSGGIGATNVLRLQSKGVSNKQWALLVL 73
QY 58 AGDILKGTALPFLMHVDIH-PLLAGVFAVLGHVPIFAKFGKGVAVATSGGVLLFYA 116
DB 74 ILDLFKGMFAVFLSKFLGLDYSLOWMVAIASILGHCHYSPFLNFGKGVSTINGSVLLI 133
QY 117 PLLEITWAVFFELYLTNKFVLSLSSML-TGIYTVIYSP--FVH--DYLLI-----VV 164
DB 134 PIESLIGLTVWFFVGVKVLKISSLASILGSGTATVLIFFVPMHPDPSVNLKVGQTQPM 193
QY 165 TLLTIFVIYHRANIKRIINTEPKV 190
DB 194 VLIFIFTLIKAGNIFNLLTCKEKKV 219

RESULT 10

YF09_HELPJ

ID YF09_HELPJ STANDARD: PRT; 220 AA.
AC Q26039;
DT 15-JUL-1999 (Rel. 38, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein HPI509.
GN HPI509.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=26695 / ATCC 700392;
MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., FitzGerald L.M., Lee N., Adams M.D., Hickey E.K.,
Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Matthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori";
RL Nature 388:539-547(1997).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE UPF0078 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AE000649; AAC08551.1; ALT_INIT.
DR TIGR: HPI509;
DR InterPro: IPR003811; DUF205.
DR Pfam: PF02660; DUF205; 1.
DR TIGRFAMs: TIGR00023; DUF205; 1.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 11 31 POTENTIAL.
FT TRANSMEM 34 54 POTENTIAL.
FT TRANSMEM 70 90 POTENTIAL.
FT TRANSMEM 128 148 POTENTIAL.
FT TRANSMEM 153 173 POTENTIAL.
SQ SEQUENCE 220 AA; 23804 MW; E72004BFF4AB882B CRC64;

Query Match 23.88; Score 228; DB 1; Length 220;
Best Local Similarity 31.68; Pred. No. 3.4e-10;
Matches 65; Conservative 38; Mismatches 83; Indels 20; Gaps 6;

QY 5 LLIIILYLGISGPSGLIVGKLAGIDIREHSGNLTGATNAPRTLGVRAGS-----VVI 57
DB 14 IFTLLGLIGIPGYALMKIFYGMDITKIGSGGIGATNVLRLQSKGVSNKQWALLVL 73
QY 58 AGDILKGTALPFLMHVDIH-PLLAGVFAVLGHVPIFAKFGKGVAVATSGGVLLFYA 116
DB 74 ILDLFKGMFAVFLSKFLGLDYSLOWMVAIASILGHCHYSPFLNFGKGVSTINGSVLLI 133
QY 117 PLLEITWAVFFELYLTNKFVLSLSSML-TGIYTVIYSP--FVH--DYLLI-----VV 164
DB 134 PIESLIGLTVWFFVGVKVLKISSLASILGSGTATVLIFFVPMHPDPSVNLKVGQTQPM 193
QY 165 TLLTIFVIYHRANIKRIINTEPKV 190
DB 194 VLIFIFTLIKAGNIFNLLTCKEKKV 219

RESULT 11

```

QY 114 FYAPLFIETWAVFFIFLILYLTKEVSLSSMLTGIYVIYSFFVHDTYLLIIVTLLTTFVIY 173
    || : || :
Db 272 -YATAYMLTDTVYFGV-----LAIAVPVSVIVSL 300
QY 174 RHRANIKRI 182
    ||| : ||
Db 301 LHRAGVIRL 309
    ||| : ||

RESULT 12
COX1_PONPA
ID COX1_PONPA STANDARD; PRT: 514 AA.
AC F92692;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cytochrome c oxidase polypeptide I (EC 1.9.3.1).
GN MICO1 OR COI.
OG Pongo pygmaeus abelii (Sumatran orangutan).
OC Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pongo.
OC NCBI_TaxID=9601;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=YN93-312;
RX MEDLINE=97032590; PubMed=8875856;
RA Xu X, Arnason U.;
RT "The mitochondrial DNA molecule of Sumatran orangutan and a molecular
RT proposal for two (Bornean and Sumatran) species of orangutan.";
RL J. Mol. Evol. 43:431-437(1996).
CC -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATOR
CC 3 CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1
CC FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO 1 IS THE
CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A
CC AND COPPER B.
CC -!- CATALYTIC ACTIVITY: 4 ferrocycytochrome c + O(2) = 4 ferrocycytochrome
CC c + 2 H(2)O.
CC -!- PATHWAY: Respiratory chain; terminal step.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.
CC -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collabora
CC between the Swiss Institute of Bioinformatics and the EMBL outstatio
CC the European Bioinformatics Institute. There are no restrictions on th
CC use by non-profit institutions as long as its content is in conforman
CC modified and this statement is not removed. Usage by and for commercia
CC entities requires a license agreement (See http://www.isb-sib.ch/anno
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X97707; CAA66285.1; -
CC HSP; P00396; Z0CC.
CC InterPro: IPR000883; COX1.
CC Pfam: PF00115; COX1; 1.
CC PRINTS; PR01165; CYCOXIDASE1.
CC PROSITE; PS00077; COX1; 1.
CC Oxidoreductase; Heme; Copper; Copper; Mitochondrion; Transmembrane;
CC Respiratory chain; Inner membrane.
CC METAL 61 61 IRON (HEME A) (PROBABLE).
CC METAL 240 240 COPPER B (PROBABLE).
CC METAL 244 244 COPPER B (PROBABLE).
CC METAL 290 290 COPPER B (PROBABLE).
CC METAL 291 291 COPPER B (PROBABLE).
CC METAL 376 376 IRON (HEME A3) (PROBABLE).
CC METAL 378 378 IRON (HEME A) (PROBABLE).
CC SEQUENCE 514 AA; 57009 MW; 723609BA8E293FCE CRC64;

Query Match 10.1%; Score 97; DB 1; Length 514;
Best Local similarity 24.5%; Pred. No. 2.3;

```

Matches 51: Conservative 34; Mismatches 79; Indels 44; Gaps 10;

OY 10 AYLGISPSGLIVKGLAKGIDIREHSGN-----LGTNAPRFLGVKAGSVVIAGD 60
 Db 308 ATMIATPTGVKFSWLTATL-----HGSNKRWSAAILNALGFIPLF-TVGGLTG-IVLANS 361
 OY 61 ILKGTATATLPLMHVDIHLPLAGVFAVLG---HVFPFAKFGKAVATSGGVLLFYAP 117
 Db 362 SLDIVLHDTYVVVAHFYVLSMGAVFALMGFIHWFPLFSGYTLDTQ-----YAK 411
 OY 118 LLFTMTAVVFIFLYLT-----KFVLSLSM-----LGTIVTVYVFFVHDYTLIVVTL 167
 Db 412 IHFTIM-----PIGVNLTFPFOHFLGSLGMPRRYSYDPDAYTTWNILSAGSFI SLTAVML 467
 OY 168 TIFVYHRANIKRI--INKTEPKVKWL 193
 Db 468 MIFMIWEAFASKRVPMPVOPSTSLLEWL 495

RESULT 13
 COX1_PIG * STANDARD; PRT; 514 AA.
 ID COX1_PIG * STANDARD; PRT; 514 AA.
 DT 079876; Q9TOR4;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Cytochrome c oxidase polypeptide I (EC 1.9.3.1).
 GN MT001 OR COI.
 OS Sus scrofa (Pig).
 OC Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OX NCBI_TaxID=9623;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98404150; PubMed=9732457;
 RA Ursing B.M., Arnason U.
 RT "The complete mitochondrial DNA sequence of the pig (Sus scrofa).";
 RL J. Mol. Evol. 47:302-306(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Landrace; TISSUE=Heart;
 RX MEDLINE=99365306; PubMed=10433971;
 RA Lin C.S., Sun Y.L., Liu C.Y., Yang P.C., Chang L.C., Cheng I.C.,
 MAO S.J.T., Huang M.C.;
 RT "Complete nucleotide sequence of pig (Sus scrofa) mitochondrial genome
 and dating evolutionary divergence within artiodactyla.";
 RL Gene 236:107-114(1999).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Meishan, and Swedish wild boar;
 RA Kijas J.M.H., Andersson L.;
 RT "A phylogenetic study of the origin of the domestic pig estimated from
 the near complete mtDNA genome.";
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
 CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
 CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
 CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
 AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
 AND COPPER B.
 CC -1- CATALYTIC ACTIVITY: 4 ferrocyclochrome c + O(2) = 4 ferricytochrome
 c + 2 H(2)O.
 CC -1- PATHWAY: Respiratory chain; terminal step.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
 INNER MEMBRANE. CONTAINS 12 POTENTIAL TRANSMEMBRANE DOMAINS.
 CC -1- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See http://www.isb-sib.ch/announce/
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AJ002189; CA05231.1; -
 DR EMBL; AF034253; AAG34187.1; -
 DR EMBL; AF104203; AAC28218.1; -
 DR EMBL; AF104200; AAC28179.1; -
 DR HSP396; Z0CC.
 DR InterPro; IPR000883; COX1.
 DR Pfam; PF00115; COX1; 1.
 DR PRINTS; PR01165; CYCOXIDASE1.
 DR PROSITE; PS00077; COX1; 1.
 KW Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;
 KW Respiratory chain; Inner membrane.
 FT METAL 61 61 IRON (HEME A) (PROBABLE).
 FT METAL 240 240 COPPER B (PROBABLE).
 FT METAL 244 244 COPPER B (PROBABLE).
 FT METAL 290 290 COPPER B (PROBABLE).
 FT METAL 291 291 COPPER B (PROBABLE).
 FT METAL 376 376 IRON (HEME A3) (PROBABLE).
 FT METAL 378 378 IRON (HEME A) (PROBABLE).
 FT CONFLICT 49 50 GD -> PH (IN REF. 1).
 SQ SEQUENCE 514 AA; 56958 MW; 6B5008565248CF3A CRC64;
 Query Match 9.8%; Score 92; DB 1; Length 514;
 Best Local Similarity 22.1%; Pred. No. 5.4;
 Matches 45; Conservative 38; Mismatches 85; Indels 36; Gaps 8;
 OY 10 AYLGISPSGLIVKGLAKGIDIREHSGNIGATNAFR-----TLGVKAGSVVIAGD 60
 Db 308 ATMIATPTGVKFSWLTATL-----HG-GNIKWPAMLWALGFIPLF-TVGGLTG-IVLANS 361
 OY 61 ILKGTATATLPLMHVDIHLPLAGVFAVLG---HVFPFAKFGKAVATSGGVLLFYAP 117
 Db 362 SLDIVLHDTYVVVAHFYVLSMGAVFALMGFIHWFPLFSGYTLNQ-----WAK 411
 OY 118 LLFTMTAVVFIFLYLTGFVLSLSM-----LGTIVTVYVFFVHDYTLIVVTLTIFY 171
 Db 412 IHFVIMFVGVMNTFFPOHFLGSLGMPRRYSYDPDAYTANWTISSMGFSISLTAVMLIFI 471
 OY 172 IYRHRANIKRI--INKTEPKVKWL 193
 Db 472 IWEAFASKREVSAVELTSTNLEWL 495

RESULT 14
 NKCL_MANSE STANDARD; PRT; 1060 AA.
 ID NKCL_MANSE
 AC Q25479;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Bumetanide-sensitive sodium-(Potassium)-chloride cotransporter
 (NA-K-CL symporter).
 OS Manduca sexta (Tobacco hawkmoth) (Tobacco hornworm).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata;
 OC Ditrysia; Sphingioidea; Sphingidae; Manduca;
 OX NCBI_TaxID=7130;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Malpighian tubules;
 RX MEDLINE=95035837; PubMed=7550244;
 RA Reagan J.D.;
 RT "Molecular cloning of a putative Na(+)-K(+)-2Cl-cotransporter from
 the Malpighian tubules of the tobacco hornworm, Manduca sexta.";
 RL Insect Biochem. Mol. Biol. 25:875-880(1995).
 CC -1- FUNCTION: ELECTRICALLY SILENT TRANSPORTER SYSTEM WHICH IS A
 MEDIATOR OF SODIUM AND CHLORIDE REABSORPTION. PLAYS A VITAL ROLE
 IN THE REGULATION OF IONIC BALANCE AND CELL VOLUME.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- SIMILARITY: BELONGS TO THE SLC12A FAMILY OF TRANSPORTERS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: U17344; AAA75600.1; --
 DR InterPro: IPR002293; AA/rel_pmaseel.
 DR InterPro: IPR004842; KCL_cotransport.
 DR InterPro: IPR002443; NakCl_transprter.
 DR PRINTS: PR01207; NAKCLTRNSPRT.
 DR TIGRFAMs: TIGR00930; 2a30; 1.
 KW Transport; Transmembrane; Glycoprotein.
 FT DOMAIN 1 122 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 123 143 POTENTIAL.
 FT DOMAIN 144 153 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 154 174 POTENTIAL.
 FT DOMAIN 175 197 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 198 218 POTENTIAL.
 FT DOMAIN 219 249 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 250 270 POTENTIAL.
 FT DOMAIN 271 275 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 276 296 POTENTIAL.
 FT DOMAIN 297 331 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 332 352 POTENTIAL.
 FT DOMAIN 353 367 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 368 388 POTENTIAL.
 FT DOMAIN 389 431 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 432 452 POTENTIAL.
 FT DOMAIN 453 497 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 498 518 POTENTIAL.
 FT DOMAIN 519 562 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 563 583 POTENTIAL.
 FT DOMAIN 584 642 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 643 663 POTENTIAL.
 FT DOMAIN 664 881 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 882 902 POTENTIAL.
 FT DOMAIN 903 1060 CYTOPLASMIC (POTENTIAL).
 FT CARBOHYD 396 396 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 404 404 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 419 419 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 816 816 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1060 AA; 117392 MW; 4C39D1F0B645FFEF CRC64;

Query Match 9.6%; Score 92; DB 1; Length 1060;

Best Local Similarity 23.1%; Pred. No. 9.8;

Matches 57; Conservative 34; Mismatches 84; Indels 72; Gaps 12;

QY 2 LIALLI-ILAYLI-----GSI--PSGLIVGKLAKGIDIREHSGNIGATNAFTLG-- 49
 DB 368 LLALLISMSYSLWLVFAGGGALRDASGNITDLIVNGVTVDYSSVLCALNNTCEYGLH 427
 QY 50 -----VKAGSVIAGDILKGTATLPELHMVHDIDHLLAGVFAVLGHVPE---IPAK 98
 DB 428 NSYSVMQLMSANGPFYGGCWARTLSTALTNLVS---PRLIQALGV-DRITPGLIFFS 483
 QY 99 FKGGKAVATSGGVLLFYAPLLFI-----TWAVFFI----- 129
 DB 484 PYGRHCEPYRGVLTFFVSLLLFIADLNTIAPLISNEFSLASVALINCFTEHRLVRLPG 543
 QY 130 -----FLYLKFKVLSLNLGTYVYFFVHDYLLIVTLTIFVIYHRANIKRIINK 185
 DB 544 WRPTFYRNWMLSLAGL---MCVAILMLVH--WNVSLVTFAIFTLY-----LIVHY 591
 QY 186 TEPKVKW 192
 DB 592 RRPDVNW 598

RESULT 15

NAHL_HUMAN STANDARD; PRT; 815 AA.
 ID NAHL_HUMAN AC P19634;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Sodium/hydrogen exchanger 1 (Na⁺)/H⁺ exchanger 1 (NHE-1) (Na⁺/H⁺ antipporter, amiloride-sensitive) (APNH).
 DE SLC9A1 OR NHE1 OR APNH1.
 GN Homo sapiens (Human).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Kidney;
 RX MEDLINE=89106219; PubMed=2536298;
 RA Sardet C., Franchi A., Pouyssegur J.;
 RT "Molecular cloning, primary structure, and expression of the human growth factor-activatable Na⁺/H⁺ antipporter.";
 RL Cell 56:271-280(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Kidney;
 RX MEDLINE=90140739; PubMed=2154036;
 RA Sardet C., Counillon L., Franchi A., Pouyssegur J.;
 RT "Growth factors induce phosphorylation of the Na⁺/H⁺ antipporter, glycoprotein of 110 kD.";
 RL Science 247:723-726(1990).
 RN [3]
 RP REVISIONS.
 RX MEDLINE=91293066; PubMed=1712287;
 RA Tse C.-N., Ma A.I., Yang V.W., Watson A.J.M., Levine S., Montrose M.H., Potter J., Sardet C., Pouyssegur J., Donowitz M.;
 RT "Molecular cloning and expression of a cDNA encoding the rabbit ileal villus cell basolateral membrane Na⁺/H⁺ exchanger.";
 RL EMBO J. 10:1957-1967(1991).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Heart;
 RX MEDLINE=94111706; PubMed=8283968;
 RA Fliegel L., Dyck J.R., Wang H., Fong C., Haworth R.S.;
 RT "Cloning and analysis of the human myocardial Na⁺/H⁺ exchanger.";
 RL Mol. Cell. Biochem. 125:137-143(1993).
 RN [5]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20375279; PubMed=10913675;
 RA Garden O.A., Musk P., Worthington-White D.A., Dewey M.J., Rich I.N.;
 RT "Silent polymorphisms within the coding region of human sodium/hydrogen exchanger isoform-1 cDNA in peripheral blood mononuclear cells of leukemia patients: A comparison with healthy controls.";
 RL Cancer Genet. Cytogenet. 120:37-43(2000).
 CC -!- FUNCTION: INVOLVED IN PH REGULATION TO ELIMINATE ACIDS GENERATED BY ACTIVE METABOLISM OR TO COUNTER ADVERSE ENVIRONMENTAL CONDITIONS. MAJOR PROTON EXTRUDING SYSTEM DRIVEN BY THE INWARD SODIUM ION CHEMICAL GRADIENT. PLAYS AN IMPORTANT ROLE IN SIGNAL TRANSDUCTION.
 CC -!- SUBUNIT: Interacts with tescalcin.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- TISSUE SPECIFICITY: KIDNEY AND INTESTINE.
 CC -!- PTM: PHOSPHORYLATED (POSSIBLE).
 CC -!- MISCELLANEOUS: INHIBITED BY AMILORIDE AND 5-AMINO-SUBSTITUTED DERIVATIVES AND ACTIVATED IN A COOPERATIVE FASHION BY INTRACELLULAR H⁺. FULLY ACTIVE AT ACIDIC PH, THE ANTIPOPPER IS VIRTUALLY TURNED OFF AT NEUTRAL PH. IN QUIESCENT CELLS UPON GROWTH FACTOR STIMULATION, THE APPARENT AFFINITY FOR INTERNAL H⁺ IS INCREASED, RESULTING IN A PERSISTENT RISE IN CYTOPLASMIC PH.
 CC -!- SIMILARITY: BELONGS TO THE NA⁺/H⁺ EXCHANGER FAMILY.
 CC -!- CAUTION: THE NUMBER, LOCALIZATION AND DENOMINATION OF HYDROPHOBIC DOMAINS IN THE NA⁺/H⁺ EXCHANGERS VARY AMONG AUTHORS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
 CC or send an email to license@isb-sib.ch.
 CC -----

DR EMBL: M81768; AAB59460.1; ALT_SEQ.
 DR EMBL: S68616; AAC60606.1; -
 DR EMBL: AF141350; AAF21350.1; -
 DR EMBL: AF141351; AAF21351.1; -
 DR EMBL: AF141352; AAF21352.1; -
 DR EMBL: AF141353; AAF21353.1; -
 DR EMBL: AF141354; AAF21354.1; -
 DR EMBL: AF141355; AAF21355.1; -
 DR EMBL: AF141356; AAF21356.1; -
 DR EMBL: AF141357; AAF21357.1; -
 DR EMBL: AF141358; AAF21358.1; -
 DR EMBL: AF141359; AAF21359.1; -
 DR EMBL: AF146430; AAF25592.1; -
 DR EMBL: AF146431; AAF25593.1; -
 DR EMBL: AF146432; AAF25594.1; -
 DR EMBL: AF146433; AAF25595.1; -
 DR EMBL: AF146434; AAF25596.1; -
 DR EMBL: AF146435; AAF25597.1; -
 DR EMBL: AF146436; AAF25598.1; -
 DR EMBL: AF146437; AAF25599.1; -
 DR EMBL: AF146438; AAF25600.1; -
 DR EMBL: AF146439; AAF25601.1; -
 DR PIR: A31311; A31311.
 DR Genew: HGNC:11071; SLC9A1.
 DR MIN: 107310; -
 DR InterPro: IPR000676; NaH_Exchng.
 DR InterPro: IPR004709; NaH_exchang3.
 DR Pfam: PF00999; Na_H_Exchange; 1.
 DR PRINTS: PR01084; NAHEXCHNGR.
 DR TIGRFAMs: TIGR00840; b_cpai.1.
 DR Transmembrane: Glycoprotein; Transport; Antiport; Sodium transport;
 KW Multigene family; Phosphorylation.

FT DOMAIN 1 15 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 16 35 M1 (POTENTIAL).
 FT DOMAIN 36 107 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 108 127 M2 (POTENTIAL).
 FT DOMAIN 128 129 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 130 149 M3 (POTENTIAL).
 FT DOMAIN 150 154 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 155 174 M4 (POTENTIAL).
 FT DOMAIN 175 191 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 192 211 M5 (POTENTIAL).
 FT DOMAIN 212 227 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 228 247 M5A (POTENTIAL).
 FT DOMAIN 248 256 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 257 276 M5B (POTENTIAL).
 FT DOMAIN 277 294 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 295 315 M6 (POTENTIAL).
 FT DOMAIN 316 338 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 339 358 M7 (POTENTIAL).
 FT DOMAIN 359 386 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 387 406 M8 (POTENTIAL).
 FT DOMAIN 407 410 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 411 430 M9 (POTENTIAL).
 FT DOMAIN 431 480 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 481 500 M10 (POTENTIAL).
 FT DOMAIN 501 815 CYTOPLASMIC (POTENTIAL).
 FT CARBOHYD 75 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 370 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 815 AA; 90763 MW; 02EC748C79DF6526 CRC64;

Query Match 9.58; Score 91.5; DB 1; Length 815;
 Best Local Similarity 20.78; Pred. No. 8.6;
 Matches 58; Conservative 50; Mismatches 79; Indels 93; Gaps 12;
 QY 3 IALLIILA-----YLIGSIPS-----GLIVGNLAKGID----- 30

Db 105 ISLWILLACLMMKIGFHVPTISSIVPESCLLIIVVGLLVGLIKGVGETPPPLQSDVFLF 164
 Qy 31 -----IREHG-----SGNLGATNAPRTLGVRAGSVVIAG----- 59
 Db 165 LLPPIILDAGYFLPLROFTENLGTILIFAVVGTLWNAFFLGLMAYAVCLVGEQINNI 224
 Qy 60 --DILKGTALATLPFL-----MHVD--IHPLLAG-----VFVGLGHVFIKPK 100
 Db 225 LDNLFGSIISAVDPVAVLAVFEEIHNELHLVFGESLLNDVAVTVVLYHLFEFANYE 284
 Qy 101 GSKAVATSGGVLLFVAPLLFITMVAVFP--IFLYLTRKVSLSMMLTGITYTYISFFVHDT 158
 Db 285 HVGIVDFLGSFSEFFVALGVGVYGVIAAFTSRFTSHRVEIPLFVFLYSTMAYLS 344
 Qy 159 ----YLLIVVTLLTIFVIYRH--RANIKRIINKTEPKVKW 192
 Db 345 AELFHLSCIMALIASGVVMRYPVEANIS---HKSHTTIKY 381

Search completed: December 26, 2002, 00:54:17
 Job time : 140.382 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 26, 2002, 00:00:26 : Search time 623.209 Seconds
(without alignments)
63.810 Million cell updates/sec

Title: US-10-068-080-3
Perfect score: 960
Sequence: 1 MLTALLIILAYLIGSIPSGL.....RHRANIKRIINKTEPKVKWL 193

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriaph:*
- 17: sp_archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	568.5	59.2	198	16 Q92C68	Q92C68 listeria in
2	566.5	59.0	198	16 Q8Y7J3	Q8Y7J3 listeria mo
3	464.5	48.4	202	16 Q99UC5	Q99UC5 staphylococ
4	453	47.2	213	16 Q9A070	Q9A070 streptococ
5	445	46.4	213	16 Q9CGW4	Q9CGW4 lactococcus
6	430	44.8	194	16 Q8R9Y9	Q8R9Y9 fusobacteri
7	385.5	40.2	168	2 Q9X972	Q9X972 streptococ
8	373.5	38.9	198	16 Q8R9J2	Q8R9J2 thermoanaer
9	339.5	35.4	226	16 Q8Y2G8	Q8Y2G8 anabaena sp
10	333	34.7	205	16 Q8UFU1	Q8UFU1 agrobacteri
11	322	33.5	206	16 Q9KCD3	Q9KCD3 bacillus ha
12	309.5	32.2	196	16 Q9X1F9	Q9X1F9 thermotoga
13	301.5	31.4	199	2 Q9ZAF3	Q9ZAF3 thermus the
14	301.5	31.4	203	16 Q9ZQL7	Q9ZQL7 rhizobium m
15	301	31.4	201	16 Q8YC64	Q8YC64 brucella me
16	299	31.1	208	16 Q9KUJ7	Q9KUJ7 vibrio chol

17	296	30.8	198	16 Q9RSV1	Q9RSV1 deinococcus
18	292.5	30.5	195	16 Q9BM84	Q9BM84 rhizobium l
19	286	29.8	216	16 Q8Z167	Q8Z167 yersinia pe
20	284.5	29.6	218	16 Q9A5K1	Q9A5K1 caulobacter
21	271.5	28.3	198	16 Q97G69	Q97G69 clostridium
22	270.5	28.2	257	16 Q9PQ85	Q9PQ85 ureaplasma
23	270	28.1	207	16 Q8XWC8	Q8XWC8 ralstonia s
24	266.5	27.8	201	16 Q9CKC7	Q9CKC7 pasteurella
25	263.5	27.4	202	16 Q9FIE4	Q9FIE4 campylobact
26	261.5	27.2	224	16 Q9QR6	Q9QR6 mycoplasma
27	257.5	26.8	203	16 Q8XGX7	Q8XGX7 salmonella
28	257	26.8	200	16 Q9JZG9	Q9JZG9 neisseria m
29	257	26.8	200	16 Q9JUL4	Q9JUL4 neisseria m
30	252	26.2	189	16 Q915V6	Q915V6 pseudomonas
31	198.5	20.7	203	16 Q9X1O9	Q9X1O9 thermotoga
32	165	17.2	195	16 Q9RS57	Q9RS57 deinococcus
33	141	14.7	559	16 Q9RW79	Q9RW79 deinococcus
34	133	13.9	425	17 Q8TH23	Q8TH23 methanosarc
35	115	12.0	463	16 Q34474	Q34474 bacillus su
36	110.5	11.5	421	16 Q9XOD5	Q9XOD5 thermotoga
37	108.5	11.3	243	17 Q8TQJ0	Q8TQJ0 methanosarc
38	104.5	10.9	963	16 Q8YU47	Q8YU47 anabaena sp
39	104	10.8	325	17 Q973Q6	Q973Q6 sulfolobus
40	101.5	10.6	381	17 Q9H1Z1	Q9H1Z1 thermoplasma
41	103	10.4	374	17 Q97X02	Q97X02 sulfolobus
42	103	10.4	671	17 Q8TJA8	Q8TJA8 methanosarc
43	99	10.3	477	16 Q8XP64	Q8XP64 clostridium
44	99	10.3	513	8 Q9TX97	Q9TX97 pongo pygma
45	98.5	10.3	276	17 Q9UZ69	Q9UZ69 pyrococcus

ALIGNMENTS

RESULT 1
Q92C68 ID Q92C68 PRELIMINARY; PRT; 198 AA.
AC Q92C68; 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical protein lin1323.
GN LIN1323.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Listeriaceae; Listeria.
OX NCBI_TaxID=1642;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIP 11262 / SEROVAR 6A;
RX PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
BA Quero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kref J., Kuhn M., Kunst F., Kurapatk.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Medjari H.,
RA Nordsiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
RA Rammel B., Rose M., Schluetter T., Simoes N., Tierrez A.,
Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT *Comparative genomics of Listeria species.*;
RL Science 294:849-852(2001).
EMBL; AL596168; CAC96554.1;
DR ListList: LIN01323;
DR InterPro: IPR003811; DUF205.
DR Pfam: PF02660; DUF205; 1.
DR TIGRFAMs: TIGR00023; DUF205; 1.
DR Hypothetical protein; Complete proteome.
SQ SEQUENCE 198 AA; 21632 MW; B161D1055B203406 CRC64;
Query Match 59.2%; Score 568.5; DB 16; Length 198;

	Best Local Similarity	57.7%; Pred.	No. 2.3e-40;	Gaps	1:
	Matches	112; Conservative	35; Mismatches	44; Indels	3;
Qy	3	IALLIILAYLIGSIPSGLVGKLAGIDIREHGSNGLGNATNAFRTLGVKGAGSVVIAGDIL	62		
Dd	5	LILLSLAYVIGSIPSGWLWICKYFKKDIREFGSNGLGNATNSFRLVGKAGSIVTVMDIL	64		
Qy	63	KGTALPALPFLMHVDIHDP---LLAGVFVLAVGHVPFIFAKFVKGGKA VATSGGVLFFYAPLL	119		
Dd	65	KGTVALTLPFPFNLNHNHHFWLLTGAFAILGHSPFLFAGRGRGA VTSAGVILAYAPLL	124		
Qy	120	FITWVA VF FLYLT KFSVLS SML TGI VT YVS FF VHDTY LL I LVVT LTI FT VIYHRANI	179		
Dd	125	FVAALVFEPLTKISKYSVSSMTCALAALIISPFMGDWIIILIACIALFVINRH RANI	184		
Qy	180	KRIINKTEPKVKWL	193		
Dd	185	TRIRNGEPEKKWM	198		
RESULT 2					
Q8Y7J3	ID	PRELIMINARY;	PRT;	198 AA.	
AC	QB87J3;				
DT	01-MAR-2002	(TrEMBLrel. 20, Created)			
DT	01-MAR-2002	(TrEMBurel. 20, Last sequence update)			
DT	01-JUN-2002	(TrEMBurel. 21, Last annotation update)			
DE	Hypothetical protein Imol284.				
DN	LMO1284.				
OS	Listeria monocytogenes.				
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;				
CC	Listeriaceae; Listeria.				
OX	NCB_I_TaxID=1639;				
RN	[[]]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=EGD-E / SEROVAR 1/2A;				
RX	MEDLINE=N1537279; PubMed=11679669;				
RA	Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,				
RA	Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,				
RA	Charbit A., Chetoui F., Couve E., de Daruvar A., Dehoux P.,				
RA	Dommann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,				
RA	Elian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,				
RA	Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,				
RA	Jones L.-M., Kastir U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,				
RA	Madueno E., Maicounan A., Mata Vicente J., Ng E., Nedjari H.,				
RA	Nordsiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,				
RA	Rommel B., Rose M., Schlueter T., Simoes N., Tierrez A.,				
RA	Vazquez-Boland J.-A., Voss H., Wehlant J., Cozzart P.;				
RT	"Comparative genomics of Listeria species.";				
RL	Science 294:849-852(2001).				
DR	EMBL; AL591978; CAC99362.1; -				
DR	ListiliList; LMO01284; -				
DR	InterPro; IPRO03811; DUF205.				
DR	Pfam; PF02660; DUF205; 1.				
DR	TIGRFAMs; TIGR00023; DUF205; 1.				
KW	Hypothetical protein; Complete proteome.				
SQ	SEQUENCE	198 AA; 21599 MW; 7807B5406DF05CD1 CRC64;			
Query Match		Score 566.5; DB 16; Length 198;			
Best Local Similarity		58.8%; Pred.	No. 3.4e-40;		
Matches	114; Conservative	33; Mismatches	44; Indels	3;	Gaps
Qy	3	IALLIILAYLIGSIPSGLVGKLAGIDIREHGSNGLGNATNAFRTLGVKGAGSVVIAGDIL	62		
Dd	5	LILLSLAYVIGSIPSGWLWICKYFKKDIREFGSNGLGNATNSFRLVGKAGSIVTVMDIL	64		
Qy	63	KGTALPALPFL--MHVDIH-P-LLAGVFVLAVGHVPFIFAKFVKGGKA VTSAGGVLFFYAPLL	119		
Dd	65	KGTVALTLPFPFNLNHNHHFWLLTGAFAILGHSPFLFAGRGRGA VTSAGVILAYAPLL	124		
Qy	120	FITWVA VF FLYLT KFSVLS SML TGI VT YVS FF VHDTY LL I LVVT LTI FT VIYHRANI	179		
Dd	125	FVAALVFEPLTKISKYSVSSMTCALAALIISLFPMGDWIILLIACIALFVINRH RANI	184		

OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC	Streptococcaceae; Streptococcus.
OX	NCBI_TaxID=1314;
OX	{1}
RN	SEQUENCE FROM N.A.
RP	STRAIN-SF370 / ATCC 700294 / SEROTYPE M1;
RC	MEDLINE=21192684; PubMed=11296296;
RC	Perretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA	Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA	Qian Y., Jia H.G., Najaf F.Z., Ren Q., Zhu H., Song L., White J.,
RA	Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT	*Complete genome sequence of an M1 strain of <i>Streptococcus pyogenes</i> .*;
RT	Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
RL	EMBL: AE006539; AAK33824.1; -
RL	InterPro: IPR003811; DUF205.
DR	Pfam: PF02660; DUF205; 1.
DR	TIGRFAMS: TIGR00023; DUF205; 1.
DR	Hypothetical protein; Complete proteome.
KW	SEQUENCE 213 AA; 23369 MW; 6A98812332A09766A CRC64;
SC	

Query Match	47.2%	Score 453:	DB 16:	Length 213:
Best Local Similarity	47.0%:	Pred. No. 1.3e-30:		
Matches	95:	Conservative	34:	Mismatches 63:
			Indels	10: Gaps 3:
QY	1	MLTALLIAYLGSPGLVGLKAGIDIREHGSNGLGATNFRTLGVKAGSVIAGD	60	
DB	1	MKULLFTIAYLGSPDGLWIGQYFHINLRHGSNGTGTNTFRILGVKAGTATLAI	60	
QY	61	ILKGTALATLFFLMHV-DIHPLLAGVFAVLGHVFPPIFAKFGKAVATSGCVLFYAPLL	119	
DB	61	MFKGTLSILLPIIFGMTSISIAIGFFAVLGHTEPPIFANFKGKAVATSGVLLGCFAPLY	120	
QY	120	FIMVAVFFFLYLTKEVLSLSSM--LTGIYTVIYSFVH-----DTYLLIVVTLTIF	170	
DB	121	LFTLASIFVLVLVLSMISLASVVSVAIVGLSVLFTFFPAIHFLLPNDYFLTIVILLAFI	180	
QY	171	VYHRHRANIKRIINKTEPKVKW	192	
DB	181	IIRHKDNISRIKHHTENLIPW	202	

RESULT 5					
Q9CGW4	PRELIMINARY;	PRT;	213 AA.		
AC	Q9CGW4				
DT	01-JUN-2001 (TrEMBLrel. 17, Created)				
DT	01-JUN-2001 (TrEMBLrel. 17, Last sequence update)				
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)				
DE	Hypothetical protein ykac.				
DE	YKAC OR LI0978.				
GN	Lactococcus lactis (subsp. lactis) (Streptococcus lactis).				
OS	Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;				
OC	Streptococcaceae; Lactococcus.				
OX	NCBI TaxID-1360.				

OX RN
RN RP
RC STRAIN=IL1403;
RX MEDLINE=21235186; PubMed=11337471;
RA Bolotin A., Wincker P., Mauger S., Jaillon O., Malmarm K.,
RA Weisenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium *Lactococcus*
RT lactis ssp. *lactis* IL1403.";
RL Genome Res. 11:731-753(2001).
DR EMBL: AE006332; AAK05076.1; -.
DR InterPro: IPR003811; DUF205.
DR Pfam: PF02660; DUF205; 1.
DR TIGRFAMS: TIGR00023; DUF205; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 213 AA; 23362 MW; 967259F000F31AE3 CRC64;

Query Match 46.4%; Score 445; DB 16; Length 213;
Best Local Similarity 45.5%; Pred. No. 6.1e-30;
Matches 91; Conservative 39; Mismatches 60; Indels 10; Gaps 2;

QY	1	MLIALLIILAYLIGSPGLVIGKLAGIDIREHSGSNIAGTNAFRTLGVKAGSVVIAGD	60
		: : : : : : : : : : : : :	
Db	2	LTIIILLIASYLLGAIPFGLWICKIFFKKNLHDYSGNGTGTNTFRILGVKAGISVFAFD	61
QY	61	ILKGTATLALPFLLMHDV-THPLLGAVFANVLGVFPFAKFKGKAVATSGGVLLLFYAPILL	119
		: : : : : : : : : :	
Db	62	LLKAGTATLPLFHHNGVSPSLFGLLAVIGHTFSIDRFKGGKAVATSGAVILGESPLF	121
QY	120	FITMVAVFFFLVTLFVSLSSMLTGTYTVI-----YSFFVHDVTLIIIVVTLTLTF	170
		: : : : : : : : : : : :	
Db	122	LIVLVVFIIVLFWSLSSVIGAVFALLGLILFPISGFIILTSYDLLFSIIIFVLAI	181
QY	171	VYVHRANIKRIINKTEPKV	190
		: : : : : : : : :	
Db	182	IILRHRTNLKRIKNHGESLV	201
RESULT	6		
ID	Q8RFY9	PRELIMINARY;	PRT: 194 AA.
AC	Q8RFY9		
DT	01-JUN-2002	(TrEMBLrel. 21, Created)	
DT	01-JUN-2002	(TrEMBLrel. 21, Last sequence update)	
DE	01-JUN-2002	(TrEMBLrel. 21, Last annotation update)	
DE	DE	Hypothetical membrane-spanning protein FN0537.	
GN	FN0537		
OS	Fusobacterium nucleatum (subsp. nucleatum).		
OC	Bacteria; Fusobacteria; Fusobacterium.		
OX	NCBI_TaxID=76856;		
ON	[1]		
RC	SEQUENCE FROM N.A.		
RP	SPRAIN-ATCC 25586;		
RX	MEDLINE=21886394; PubMed=11889109;		
RA	Kapatral V., Anderson I., Ivanova N., Renzik G., Los T., Lykidis A.,		
RA	Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,		
RA	Vasileva O., Chu L., Kogan Y., Chaga O., Goltsman E., Bernal A.,		
RA	Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,		
RA	Fonstein M., Kyrpides N., Overbeek R.;		
RT	"Genome sequence and analysis of the oral bacterium Fusobacterium		
RT	nucleatum strain ATCC 25586";		
RL	J. Bacteriol. 184:2005-2018(2002).		
RL	EMBL; AE010565; AAL94733.1;		
SW	Complete proteome; Hypothetical protein.		
SO	SEQUENCE 194 AA; 21343 MW; DE327E0AE8B35915 CRC64;		

Query Match	44.8%;	Score 430;	DB 16;	Length 194;
Best Local Similarity	44.4%;	Pred. No. 1e-28;		
Matches 87: Conservative	41;	Mismatches 62;	Indels	6; Gaps 4;

[illegible]

RESULT 7	
Q9X972	
ID	Q9X972
AC	Q9X972; PRELIMINARY; PRT; 168 AA.
DT	01-NOV-1999 (TREMBlrel. 12, Created)
DT	01-NOV-1999 (TREMBlrel. 12, Last sequence update)

01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE Hypothetical 17.9 kDa protein (Fragment).
 OS Streptococcus gordonii.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
 OC Streptococcaceae; Streptococcus.
 OX NCBI_TaxID=1302;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CH1.
 RA Vriesema A.J., Dankert J., Zaat S.A.;
 RT "Isolation and characterization of promoter regions from Streptococcus
 gordonii CH1";
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ236899; CAB40549.1;
 DR InterPro: IPR003811; DUF205.
 DR Pfam: PF02660; DUF205; 1.
 KW Hypothetical protein.
 FT NON_TER 168 168
 SQ SEQUENCE 168 AA; 17921 MW; A07262BD799A478A CRC64;
 Query Match 40.2%; Score 385.5; DB 2; Length 168;
 Best Local Similarity 51.7%; Pred. No. 4.9e-25;
 Matches 75; Conservative 30; Mismatches 39; Indels 1; Gaps 1;
 OY 7 IILAYLIGSIPGLVIGKLGKIDIREHSGNLTGATNAPRTGLGVKAGSVVIAGDILKGTLL 66
 Db 8 LILAYLIGSIPGLVIGKLGKIDIREHSGNLTGATNAPRTGLGVKAGSVVIAGDILKGTLL 67
 OY 67 ATALPFLMVD-IHPLLAGVFAVLGHVFFIFAKFGKAVATSGGVLLFYAPLLFTMVA 125
 Db 68 ATLLPLFLHNGISPMIFGLAVLGHVFFIFAKFGKAVATSGGVLLFYAPLLFTMVA 127
 OY 126 VFIFLYLTKFVSLSSMLTGIYTVI 150
 Db 128 IFIVLYLGSMLASIVVAGFAII 152
 RESULT 8
 OY 8 Q8R9J2 PRELIMINARY; PRT; 198 AA.
 AC Q8R9J2
 DT 01-JUN-2002 (Tremblrel. 21, Created)
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
 DE Hypothetical protein TTE1618.
 GN TTE1618.
 OS Thermoanaerobacter tengcongensis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
 OC Thermoanaerobacteriales; Thermoanaerobacteriaceae; Thermoanaerobacter.
 OX NCBI_TaxID=119072;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MB4T / JCM11007;
 RA MEDLINE-21992816; PubMed-11997336;
 RX Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
 RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
 RA Tan H., Chen R., Wang J., Yu J., Yang H.;
 RT "A complete sequence of T. tengcongensis genome";
 RL Genome Res. 12:689-700(2002).
 DR EMBL: AE013117; AAM24820.1;
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 198 AA; 21016 MW; 252C3FF7D512BF02 CRC64;
 Query Match 38.9%; Score 373.5; DB 16; Length 198;
 Best Local Similarity 43.5%; Pred. No. 5.9e-24;
 Matches 83; Conservative 33; Mismatches 74; Indels 1; Gaps 1;
 OY 1 MLIALIILAYLIGSIPGLVIGKLGKIDIREHSGNLTGATNAPRTGLGVKAGSVVIAGD 60
 Db 1 MKFVLVAVLAYLIGCINNAIYFTYTRIDIRNYGSGNAGTATNVLRLGPKAAAPVFLD 60
 OY 61 ILKGTALATPFLMHVDIHLPLAGVFAVLGHVFFIFAKFGKAVATSGGVLLFYAPLLF 120
 Db 62 LKGTALATPFLMHVDIHLPLAGVFAVLGHVFFIFAKFGKAVATSGGVLLFYAPLLF 120
 Db 67 LKGTALATPFLMHVDIHLPLAGVFAVLGHVFFIFAKFGKAVATSGGVLLFYAPLLF 120
 OY 104 AVATSGGVLLFYAPLLTWTAVFFIFLYLTKFVSLSSMLTGIYTVIYFFVHD--TYLL 161
 Db 127 SVATSLGLIILAMNQVGLTFGVFAVVVAISIVSLSSIMGAIAVSIWVLPLOQPLFYL 186
 OY 162 IVVTLITFIVYHRAIKRIINTEPKV 190
 Db 187 FGIA-GGLYVILRHRSNIERLLAGTEPKI 214
 RESULT 10
 OY 10 Q8UFU1 PRELIMINARY; PRT; 205 AA.
 AC Q8UFU1
 DT 01-JUN-2002 (Tremblrel. 21, Created)
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
 DE Hypothetical protein Atul306.
 GN ATU1306 OR AGR_C_2402.
 OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

Db 61 VLKGVAVLLGKYFTGMPGALIAVAVGCHNPIFLKFRGGKAVATSGGVVMTINPLIG 120
 OY 121 ITWVAVFFIFLYLTKFVSLSSMLTGI-YTVIYFFVHD--TYLLVTLTLLTIFVYHRAIK 179
 Db 121 LIALAIGVAVIAITRYVSLGSMGTGATFALLNIFFNSVQVLTFAIVLALLVIFQHSNI 180
 OY 180 KRIINKTEPKV 190
 Db 181 KRLINGTESKI 191
 RESULT 9
 OY 9 Q8YZG8 PRELIMINARY; PRT; 226 AA.
 AC Q8YZG8
 DT 01-MAR-2002 (Tremblrel. 20, Created)
 DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
 DE Hypothetical protein A110492.
 GN A110492.
 OS Anabaena sp. (strain PCC 7120).
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
 OX NCBI_TaxID=103690;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-21595285; PubMed-11759840;
 RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
 RA Watanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
 RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
 RA Yasuda M., Tabata S.;
 RT "Complete genomic sequence of the filamentous nitrogen-fixing
 cyanobacterium Anabaena sp. strain PCC 7120";
 RL DNA Res. 8:205-213(2001).
 DR EMBL: AP003582; BAB72450.1;
 DR InterPro: IPR003811; DUF205.
 DR Pfam: PF02660; DUF205; 1.
 DR TIGRFAMs: TIGR00023; DUF205; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 226 AA; 23842 MW; 766BD41EF34E5497 CRC64;
 Query Match 35.4%; Score 339.5; DB 16; Length 226;
 Best Local Similarity 40.2%; Pred. No. 4.9e-21;
 Matches 84; Conservative 37; Mismatches 67; Indels 21; Gaps 5;
 OY 2 LIALIILAYLIGSIPGLVIGKLGKIDIREHSGNLTGATNAPRTGLGVKAGSVVIAGDI 61
 Db 7 LCGAVLVAVLLGSEFTGYIAVKQLGIDIREVSGSGTGATNVLRLTGKGPAGVVLGDC 66
 OY 62 LKGTALATPFLMHVDIHLPLAGVFAVLGHVFFIFAKFGKAVATSGGVLLFYAPLLF 103
 Db 67 LKGTALATPFLMHVDIHLPLAGVFAVLGHVFFIFAKFGKAVATSGGVLLFYAPLLF 103
 OY 104 AVATSGGVLLFYAPLLTWTAVFFIFLYLTKFVSLSSMLTGIYTVIYFFVHD--TYLL 161
 Db 127 SVATSLGLIILAMNQVGLTFGVFAVVVAISIVSLSSIMGAIAVSIWVLPLOQPLFYL 186
 OY 162 IVVTLITFIVYHRAIKRIINTEPKV 190
 Db 187 FGIA-GGLYVILRHRSNIERLLAGTEPKI 214
 RESULT 10
 OY 10 Q8UFU1 PRELIMINARY; PRT; 205 AA.
 AC Q8UFU1
 DT 01-JUN-2002 (Tremblrel. 21, Created)
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
 DE Hypothetical protein Atul306.
 GN ATU1306 OR AGR_C_2402.
 OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

OC Rhizobiaceae: Rhizobium.
 OX NCBI_TaxID=176299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21608550; PubMed=11743193;
 RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
 RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
 RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
 RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
 RA Kutyavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
 RA Zhang S., Yoo H., Biddle P., Jung M., Romero P., Gordon D.,
 RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Krespan W., Perry M.,
 RA Zhou S., Yoo H., Biddle P., Jung M., Romero P., Perry M.,
 RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
 RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
 RA Nester E.W.,
 RA "The genome of the natural genetic engineer Agrobacterium tumefaciens
 RT C58.";
 RL Science 294:2317-2323(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21608551; PubMed=11743194;
 RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
 RA Qurrelo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
 RA Roumel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
 RA Wollam C., Allinger M., Dougherty D., Scott C., Lappas C., Markelz B.,
 RA Flanagan C., Crowell C., Gurson J., Lomo C., Sear C., Strub G.,
 RA Cielo C., Slater S.;
 RT "Genome sequence of the plant pathogen and biotechnology agent
 RT Agrobacterium tumefaciens C58.";
 RL Science 294:2323-2328(2001).
 DR EMBL: AE009092; AAL42312.1; ALT_INIT.
 DR EMBL: AE008058; AAK87097.1; -;
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 205 AA; 21678 MW; 973A5024E0CFDE46 CRC64;

Query Match 34.7%; Score 333; DB 16; Length 205;
 Best Local Similarity 38.1%; Pred. No. 1.6e-20;
 Matches 72; Conservative 43; Mismatches 74; Indels 0; Gaps 0;

QY 2 LIALILAYLIGSIPGLIVGKLAGKIDIREHSGNIGATNAPFTLGVKAGSVIAGDI 61
 DB 13 LLALALIGVLLGSIPFGLILTRMAGLGDVRKIGSGNIGATNVLRTGNKKLAATLLDA 72
 QY 62 LAGTATLALPFLMHVDIHPHLLAGVFAVLGHVFPPIFAKFGKRAVATSGVLLFYAPLLFI 121
 DB 73 LKGTAAVLVANALMGVEASLVAGFFAFLGHLFPVLGFGKGVAVYIGVLLGAAPLMML 132
 QY 122 TWAVFFIFLYLTKFVSLSSMLTGIVTVIYSFVHDITLIVVTLTIIVVHRANIKR 181
 DB 133 AFALILWATATITRYSSLSALLAMLIPVALVGLPEKTAMLVTLISVISWKKHRENI 192
 QY 182 IINKTEPKV 190
 DB 193 LMAGTESRI 201

RESULT 11
 Q9KCD3 PRELIMINARY; PRT; 206 AA.
 AC Q9KCD3
 DT 01-OCT-2000 (TRENBLrel. 15, Created)
 DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
 DE Hypothetical protein BH1639.
 GN BH1639.
 OS Bacillus halodurans.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
 OC Bacillaceae; Bacillus.
 OX NCBI_TaxID=86665;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=C-125 / JCM 9153;
 MEDLINE=20512582; PubMed=11058132;

RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
 RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
 RA Horikoshi K.;
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
 RL halodurans and genomic sequence comparison with Bacillus subtilis.";
 DR EMBL: AP001512; BAB05358.1; -;
 DR InterPro: IPR003811; DUF205.
 DR Pfam: PF02660; DUF205; 1.
 DR TIGRFAMS: TIGR00023; DUF205; 1.
 DR Hypothetical protein; Complete proteome.
 KW SEQUENCE 206 AA; 21890 MW; 6C7CA6474E4013FA CRC64;
 SQ SEQUENCE 206 AA; 21890 MW; 6C7CA6474E4013FA CRC64;

Query Match 33.5%; Score 322; DB 16; Length 206;
 Best Local Similarity 37.3%; Pred. No. 1.3e-19;
 Matches 75; Conservative 43; Mismatches 61; Indels 22; Gaps 4;

QY 5 LLILAYLIGSIPGLIVGKLAGKIDIREHSGNIGATNAPFTLGVKAGSVIAGDI 64
 DB 6 LLVIGSYLLGSVFSYIIIAKKIKKVDIRHSGNAGATNLRVLGVGPAVTLLDILKG 65
 QY 65 TLATLALPFLMHVDIHP-----LLAGVFAVLGHVFPPIFAKFGKRAVATSGVLLFYAP 117
 DB 66 VIATV-----VVTYQLTPDGDGWFMAAAGAAIIGHNWPYIYFGKGVATTVIGVLASLVP 121
 QY 118 LLFI-----TWAVFFIFLYLTKFVSLSSMLTGIVTVIYSFVHDITLIVVTLTI 169
 DB 122 LAALVIAVIAIGSVITVTVYVSLGSLFVTLTALLVAVLSOMFGYVPVYIYTLIIIVAIL- 180
 QY 170 FVYIHRANIKRIINKTEPKV 190
 DB 181 --WHRHSRIORLLSGSTENKL 199

RESULT 12
 Q9X1F9 PRELIMINARY; PRT; 196 AA.
 AC Q9X1F9
 DT 01-NOV-1999 (TRENBLrel. 12, Created)
 DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
 DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
 DE Hypothetical protein TM1447.
 GN TM1447.
 OS Thermotoga maritima.
 OC Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
 OX NCBI_TaxID=23336;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN=MSB8 / DSM 3109;
 RX MEDLINE=99287316; PubMed=10360571;
 RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
 RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
 RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
 RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
 RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
 RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
 RT "Evidence for lateral gene transfer between Archaea and Bacteria from
 RT genome sequence of Thermotoga maritima.";
 RL Nature 399:323-329(1999).
 DR EMBL: AF001796; AAD36515.1; -;
 DR TIGR: TM1447; -;
 DR InterPro: IPR003811; DUF205.
 DR Pfam: PF02660; DUF205; 1.
 DR TIGRFAMS: TIGR00023; DUF205; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 196 AA; 21342 MW; 5AD53DD5C502AB90 CRC64;

Query Match 32.2%; Score 309.5; DB 16; Length 196;
 Best Local Similarity 37.6%; Pred. No. 1.4e-18;
 Matches 71; Conservative 34; Mismatches 79; Indels 5; Gaps 3;

QY 5 LLILAYLIGSIPGLIVGKLAGKIDIREHSGNIGATNAPFTLGVKAGSVIAGDI 64
 DB 6 LLVIGSYLLGSVFSYIIIAKKIKKVDIRHSGNAGATNLRVLGVGPAVTLLDILKG 65

Db 5 LFPILGYFIGSIPFSLIPKWLKIDVRKVGSGNVCATNAIRTTGPAVGICLLLDALKG 64
Qy 65 TLATALPFLMHVDIHL-LAGVPAVLGHVPPFAKFKGKAVATSGVLLFYAPLLFITVM 123
Db 65 FFPVFTITFTSGSKIVSTATAITVGLGHDFPFMKFKGKGVASTLGIIFCLSWPTGLVF 124
Qy 124 VAVFFIFLYLTKFVSLSSMLTGIYTVYFFV--HDTYLLIVVTLTIFVIYHRANIKR 181
Db 125 TLTLVIMVMTKYASIGSLVALYVSALLGKGYDTGMLFLI--LAVLSTLHSENIOR 182
Qy 182 IINKTEPKV 190
Db 183 LLNGTERKV 191
RESULT 13
Q9ZAF3 PRELIMINARY; PRT; 199 AA.
AC Q9ZAF3;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical 21.3 kDa protein.
OS Thermus thermophilus.
OC Bacteria; Thermus/Deinococcus group; Deinococci; Thermales;
OC Thermaceae; Thermus.
OX NCBI_TaxID=274;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HB27;
RX MEDLINE=20461243; PubMed=11004195;
RA Sanchez R., Roovers M., Glansdorff N.;
RT "Organization and expression of a thermus thermophilus arginine
RT cluster: presence of unidentified open reading frames and absence of a
RT Shine-dalgarno sequence.";
RL J. Bacteriol. 182:5911-5915(2000).
DR EMBL; Y18353; CAA77137.1; DUF205.
DR InterPro; IPR003811; DUF205.
DR Pfam; PF02660; DUF205; 1.
DR TIGRFAMs; TIGR00023; DUF205; 1.
KW Hypothetical protein.
SQ SEQUENCE 199 AA; 21331 MW; A3BFB1652F8F206B CRC64;
Query Match 31.4%; Score 301.5; DB 2; Length 199;
Best Local Similarity 38.0%; Pred. No. 6.8e-18;
Matches 71; Conservative 36; Mismatches 75; Indels 5; Gaps 4;
Qy 7 IILAYLIGSIPGLIVGKLAGIDIREHSGNLGATNAFTLGVKAGSVVIAGDILKGTPL 66
Db 8 LLLAYLFGSIPAGVLVAR--TYGVDIRKVGSGNIGATNVLRLALGWPALVAVFFDFVKGI 66
Qy 67 ATALPFLMHVDIHLPLAGV--FAVLGHVPPFAKFKGKAVATSGVLLFYAPLLFITMV 124
Db 67 AVLVARAFGLS--DMWLGGLMAYLGHVSVFLRFRGKGVATSGFTLLFLDPAALATWF 125
Qy 125 AVFFIFLYLTKFVSLSSMLTGIYTVYFFV--HDTYLLIVVTLTIFVIYHRANIKRII 183
Db 126 PIGLSVILLTRYVSGTGGVAAVFLSLALGRPLWEVATVFLMALLIFWTHRENLRKIQ 185
Qy 184 NKTEPKV 190
Db 186 ECTERRL 192
RESULT 14
Q9ZQL7 PRELIMINARY; PRT; 203 AA.
AC Q9ZQL7;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical transmembrane protein SMC01362.
GN R01302 OR SMC01362.

OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430;
RA Capella D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsperger U.,
RA Renard C., Thebault P., Vandenbol M., Weidner S., Galibert F.;
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR EMBL; AL591786; CAC45881.1;
DR InterPro; IPR003811; DUF205.
DR Pfam; PF02660; DUF205; 1.
DR TIGRFAMs; TIGR00023; DUF205; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 203 AA; 21133 MW; 69BD0C43A30017D9 CRC64;
Query Match 31.4%; Score 301.5; DB 16; Length 203;
Best Local Similarity 35.9%; Pred. No. 6.9e-18;
Matches 66; Conservative 42; Mismatches 75; Indels 1; Gaps 1;
Qy 7 IILAYLIGSIPGLIVGKLAGIDIREHSGNLGATNAFTLGVKAGSVVIAGDILKGTPL 66
Db 17 LVFVGLGSIPLGLITRMAGLDGVRKIGSGNIGATNVLTRGNRKLAAATLFLDALKGTA 76
Qy 67 ATALPFLMHVDIHLPLAGVFAVLGHVPPFAKFKGKAVATSGVLLFYAPLLFITMVAV 126
Db 77 AAIAIYAGVGA-GIAAGFAAFLGHLFPVWLSFRGKGVATVIGVLLGLMPVWVLLFAAI 135
Qy 127 FFIFLYLTKFVSLSSMLTGIYTVYFFV--HDTYLLIVVTLTIFVIYHRANIKRIINT 186
Db 136 WLAMAKITRYSLSALVATAVPIALYAGNGKVGALFAVTAIAWIKHRANIORLLSGT 195
Qy 187 EPKV 190
Db 196 ESRI 199
RESULT 15
Q8YC64 PRELIMINARY; PRT; 201 AA.
AC Q8YC64;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Putative integral membrane protein.
GN BMEI10663.
OS Brucella melitensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29459;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=16M / ATCC 23456 / BIOTYPE 1;
RX MEDLINE=20020109; PubMed=11756688;
RA DelVecchio V.G., Kapatral V., Redkar R.J., Patra G., Mijer C., Los T.,
RA Ivanova N., Anderson I., Bhattacharya A., Lykidis A., Reznik G.,
RA Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goltzman E.,
RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
RA Haselkort R., Kyrpides N., Overbeek R.;
RT "The genome sequence of the facultative intracellular pathogen
RT Brucella melitensis.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
DR EMBL; AE009702; AAL53910.1;
DR InterPro; IPR003811; DUF205.
DR Pfam; PF02660; DUF205; 1.
DR TIGRFAMs; TIGR00023; DUF205; 1.

Thu Dec 26 09:18:39 2002

KW Complete proteome.
 SQ SEQUENCE 201 AA; 20507 MW; 105CA44587BB4CA1 CRC64;
 Query Match 31.4%; Score 301; DB 16; Length 201;
 Best Local Similarity 39.1%; Pred. No. 7.5e-18;
 Matches 75; Conservative 33; Mismatches 78; Indels 6; Caps 3;
 Qy 1 MLIALIILAYLGSIPIGKLGKIDIREHSGNLGATNAFRTLGVKAGSVVIAGD 60
 Db 10 MLIGAL-IFGYVLGSIPIFGILITRLAGLDVRAIGSGNIGATNVLRTGNKKLAATLILD 68
 Qy 61 ILKGTLAT--ALPFLMHVDIHPLLAGVFAVLGHVFFIFARFKGKGVATSGGVLLFYAPL 118
 Db 69 ALKGTAAALIAAHFGQNAI---AAGFCATFGLFPVWIGFKGKGVATLVGLIGLAWA 125
 Qy 119 LFIWVAVFFIYLTFRVLSLMTGTIYTVIYSFFVHDYLLIIVVTLTIFVIYHRAN 178
 Db 126 CALVFAAAWIVTALLARYSSLSALVASLVVPIALYSRGNQALAAALFAIMTVIVFIKHRAN 185
 Qy 179 IKRIINKTEPKV 190
 Db 186 ISRLINGTESKI 197

Search completed: December 26, 2002, 01:16:20
 Job time : 624.209 secs

RESULT 2
US-09-134-001C-5239
; Sequence 5239, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AL
; TITLE OF INVENTION: EPIDERMIDIS FOR AD

.....

QY 72 FLM-----HV-----DIHPLLAGVFAVLGHVFPPIPAKFKGKRAVA-- 106
 Db 227 FVVISLGYIGNINIPSDTLKELKAKDQNGITLTTMTATKG--FGTFCKYLLGLIIVSLA 284
 QY 107 ----TSGVLL----FYAPLLFTWAVFFIFLYLTKFV-----SLSSMLTGIYTVIYS 152
 Db 285 CLTTAGGLIVSVSEFHRILPKYKVFIFVILVSFILANQGLNSVIKMSVPVLSVIYP 344
 QY 153 FFVHDYLLIVVTLTIFVIYR 174
 Db 345 -----VAITVILLILIAIR 357

RESULT 9

US-08-465-980-2
 ; Sequence 2, Application US/08465980
 ; Patent No. 5756309
 ; GENERAL INFORMATION:
 ; APPLICANT: Soppet, Daniel R.
 ; APPLICANT: Li, Yi
 ; APPLICANT: Rosen, Craig A.
 ; APPLICANT: Ruben, Steven M.
 ; TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR HPAJ70
 ; NUMBER OF SEQUENCES: 8
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN, CECCHI,
 ; ADDRESSEE: STUART & OLSTEIN
 ; STREET: 6 Becker Farm Road
 ; CITY: Roseland
 ; STATE: New Jersey
 ; COUNTRY: USA
 ; ZIP: 07068

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/465,980

FILING DATE: 06-JUN-1995
 CLASSIFICATION: 536
 ATTORNEY/AGENT INFORMATION:
 NAME: Ferraro, Gregory D.
 REGISTRATION NUMBER: 36,134
 REFERENCE/DOCKET NUMBER: 325800-446

TELECOMMUNICATION INFORMATION:
 TELEPHONE: 201-994-1700
 TELEFAX: 201-994-1744
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 320 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein

US-08-465-980-2

Query Match 8.5% Score 81.5; DB 1; Length 320;
 Best Local Similarity 23.8%; Pred. No. 0.59;
 Matches 41; Conservative 26; Mismatches 54; Indels 51; Gaps 6;

QY 8 ILAYLIGSI---PSGLIVGKLAAGIDIREHSGNMGATNAPRTLGKAGSVVIAGDILKG 64
 Db 145 IVAVVSGSLFFFPPLLIKRL-----AFCHSNVLSHSCVHODVMKL 186
 QY 65 TLATALP-----FLMHVDIHLPLAGVFAVLGHVFPPIPAKFKGKAVATS-----GG 110
 Db 187 AYADTLPNVYGLTALLVMGVDVMFISLSYFLIIRTVLQPLPSKSERAKAFGTCVSHIGV 246
 QY 111 VLLFYAPLLFTWAVFFIFLYLTKFVLSLSSMLTGIYTVIYSFFVHDYTYLLI 162
 Db 247 VLAFTVPLIGLSVWHRF-----GNSLHPIVRVV-----MGDIYLLL 282

RESULT 10

US-09-053-303-2
 ; Sequence 2, Application US/09053303
 ; Patent No. 5948890
 ; GENERAL INFORMATION:
 ; APPLICANT: Soppet, Daniel R.
 ; APPLICANT: Li, Yi
 ; APPLICANT: Rosen, Craig A.
 ; APPLICANT: Ruben, Steven M.
 ; TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR HPAJ70
 ; NUMBER OF SEQUENCES: 8
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN, CECCHI,
 ; ADDRESSEE: STUART & OLSTEIN
 ; STREET: 6 Becker Farm Road
 ; CITY: Roseland
 ; STATE: New Jersey
 ; COUNTRY: USA
 ; ZIP: 07068

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/053,303

FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/465,980

FILING DATE: 06-JUN-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Ferraro, Gregory D.
 REGISTRATION NUMBER: 36,134
 REFERENCE/DOCKET NUMBER: 325800-446

TELECOMMUNICATION INFORMATION:
 TELEPHONE: 201-994-1700
 TELEFAX: 201-994-1744
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 320 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein

US-09-053-303-2

Query Match 8.5% Score 81.5; DB 2; Length 320;
 Best Local Similarity 23.8%; Pred. No. 0.59;
 Matches 41; Conservative 26; Mismatches 54; Indels 51; Gaps 6;

QY 8 ILAYLIGSI---PSGLIVGKLAAGIDIREHSGNMGATNAPRTLGKAGSVVIAGDILKG 64
 Db 145 IVAVVSGSLFFFPPLLIKRL-----AFCHSNVLSHSCVHODVMKL 186
 QY 65 TLATALP-----FLMHVDIHLPLAGVFAVLGHVFPPIPAKFKGKAVATS-----GG 110
 Db 187 AYADTLPNVYGLTALLVMGVDVMFISLSYFLIIRTVLQPLPSKSERAKAFGTCVSHIGV 246
 QY 111 VLLFYAPLLFTWAVFFIFLYLTKFVLSLSSMLTGIYTVIYSFFVHDYTYLLI 162
 Db 247 VLAFTVPLIGLSVWHRF-----GNSLHPIVRVV-----MGDIYLLL 282

RESULT 11

US-09-605-785-527
 ; Sequence 527, Application US/09605785
 ; Patent No. 6321716
 ; GENERAL INFORMATION:
 ; APPLICANT: Xu, Jiangchun
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Mitcham, Jennifer L.
 ; APPLICANT: Harlocker, Susan L.
 ; APPLICANT: Jiang, Yuqi

```

; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedrick, Thomas S.
; APPLICANT: Carter, Darick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C16
; CURRENT APPLICATION NUMBER: US/09/605,785
; CURRENT FILING DATE: 2000-06-27
; NUMBER OF SEQ ID NOS: 835
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 527
; LENGTH: 320
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-605-785-527

Query Match      8.5%; Score 81.5; DB 4; Length 320;
Best Local Similarity 23.8%; Pred. No. 0.59;
Matches 41; Conservative 26; Mismatches 54; Indels 51; Gaps 6;

QY 8 ILAYLIGSI---PSGLIVGKLAIGDIREHSGNLGATNAPRTLGKAGSVVIAGDILKG 64
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 145 IVAVVRGSLFFPLPLLIKRL-----AFCHSNVLSHSCVHODVMKL 186

QY 65 TLATALP-----FLMHVDIHPLLAGVFAVLGHVFPFAKFGKGVKAVATS-----GG 110
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 187 AYADTLPNVYGLTAILLVMGVDVMFISLSYFLIIRTVLQLPKSKERAKAFGTCVSHIGV 246

QY 111 VLLFYAPLLFITMVAVFFFLYLTFRVSLSSMLTGIVTVIYVFFVHDYLLI 162
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 247 VLAFYVPLIGLSVVRHF-----GNSLHPVIRVW-----MGDIYLLL 282

RESULT 12
US-09-439-313-527
; Sequence 527, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang Yuqi
; APPLICANT: Reed, Steven G.
; APPLICANT: Kalos, Michael
; APPLICANT: Fanger, Gary
; APPLICANT: Retter, Mark
; APPLICANT: Solk, John
; APPLICANT: Day, Craig
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; FILE REFERENCE: 210121.427C9
; CURRENT APPLICATION NUMBER: US/09/439,313
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 575
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 527
; LENGTH: 320
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-439-313-527

Query Match      8.5%; Score 81.5; DB 4; Length 320;
Best Local Similarity 23.8%; Pred. No. 0.59;
Matches 41; Conservative 26; Mismatches 54; Indels 51; Gaps 6;

QY 8 ILAYLIGSI---PSGLIVGKLAIGDIREHSGNLGATNAPRTLGKAGSVVIAGDILKG 64
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 145 IVAVVRGSLFFPLPLLIKRL-----AFCHSNVLSHSCVHODVMKL 186

QY 65 TLATALP-----FLMHVDIHPLLAGVFAVLGHVFPFAKFGKGVKAVATS-----GG 110
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 187 AYADTLPNVYGLTAILLVMGVDVMFISLSYFLIIRTVLQLPKSKERAKAFGTCVSHIGV 246

QY 111 VLLFYAPLLFITMVAVFFFLYLTFRVSLSSMLTGIVTVIYVFFVHDYLLI 162
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 247 VLAFYVPLIGLSVVRHF-----GNSLHPVIRVW-----MGDIYLLL 282

RESULT 13
US-09-339-115-2
; Sequence 2, Application US/09339115
; Patent No. 6372891
; GENERAL INFORMATION:
; APPLICANT: Soppet, Daniel R.
; APPLICANT: Li, Yi
; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: HUMAN G-PROTEIN RECEPTOR HPAJ70
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN, CECCHI,
; STREET: 6 Becker Farm Road
; CITY: Roseland
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/09/339,115
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/053,303
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferraro, Gregory D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 325800-446
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 320 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-339-115-2

Query Match      8.5%; Score 81.5; DB 4; Length 320;
Best Local Similarity 23.8%; Pred. No. 0.59;
Matches 41; Conservative 26; Mismatches 54; Indels 51; Gaps 6;

QY 8 ILAYLIGSI---PSGLIVGKLAIGDIREHSGNLGATNAPRTLGKAGSVVIAGDILKG 64
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 145 IVAVVRGSLFFPLPLLIKRL-----AFCHSNVLSHSCVHODVMKL 186

QY 65 TLATALP-----FLMHVDIHPLLAGVFAVLGHVFPFAKFGKGVKAVATS-----GG 110
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB 187 AYADTLPNVYGLTAILLVMGVDVMFISLSYFLIIRTVLQLPKSKERAKAFGTCVSHIGV 246

```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 26, 2002, 01:16:25 ; Search time 80.8128 Seconds
(without alignments)
41.350 Million cell updates/sec

Title: US-10-068-080-3
Perfect score: 960
Sequence: 1 MLIALLLIAYLGISIPSGL.....RHRANKRIINKTEPKVKWL 193

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 109717 seqs, 17314136 residues

Total number of hits satisfying chosen parameters: 109717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	960	100.0	193	12	US-10-068-080-3
2	468	48.8	213	10	US-09-815-242-13329
3	468	48.8	213	10	US-09-815-242-13336
4	468	48.8	213	12	US-10-068-080-1
5	465.5	48.5	202	10	US-09-815-242-5305
6	465.5	48.5	202	10	US-09-815-242-12137
7	464.5	48.4	202	10	US-09-823-246-2
8	261.5	27.2	205	10	US-09-815-242-11713
9	257.5	26.8	203	10	US-09-815-242-13757
10	255	26.6	199	10	US-09-815-242-10997
11	252	26.2	189	10	US-09-815-242-11799
12	250.5	26.1	200	10	US-09-815-242-10303
13	230	24.0	220	10	US-09-815-242-11614
14	228	23.8	262	10	US-09-815-242-11453
15	93.5	9.7	299	10	US-09-393-634-35
16	92.5	9.6	451	10	US-09-815-242-10547
17	90	9.4	513	10	US-09-098-079-15
18	88	9.2	467	10	US-09-767-041-23
19	85	8.9	933	10	US-09-815-242-11817

20	82.5	8.6	822	10	US-09-824-734-3	Sequence 3, Appli
21	81.5	8.5	320	9	US-09-968-033C-4	Sequence 4, Appli
22	81.5	8.5	320	9	US-10-012-896-527	Sequence 527, App
23	81.5	8.5	320	9	US-09-895-793-527	Sequence 527, App
24	81.5	8.5	320	9	US-09-895-814-527	Sequence 527, App
25	81.5	8.5	320	10	US-09-759-143-527	Sequence 527, App
26	81.5	8.5	320	10	US-09-730-018-7	Sequence 7, Appli
27	81.5	8.5	320	10	US-09-780-669-527	Sequence 527, App
28	81.5	8.5	320	10	US-09-822-827-527	Sequence 527, App
29	81.5	8.5	320	10	US-09-886-055-83	Sequence 83, Appli
30	81.5	8.5	320	12	US-10-079-719-2	Sequence 2, Appli
31	81.5	8.5	320	10	US-09-789-919-66	Sequence 66, Appli
32	79.5	8.3	320	9	US-09-968-033C-2	Sequence 2, Appli
33	79	8.2	439	10	US-09-815-242-5660	Sequence 5660, Ap
34	79	8.2	484	10	US-09-815-242-12272	Sequence 12272, A
35	79	8.2	650	10	US-09-815-242-13341	Sequence 13341, A
36	78.5	8.2	430	10	US-09-815-242-5678	Sequence 5678, Ap
37	78.5	8.2	468	10	US-09-860-232A-8	Sequence 8, Appli
38	77.5	8.1	273	10	US-09-939-980-395	Sequence 395, App
39	77.5	8.1	391	10	US-09-815-242-11566	Sequence 11566, A
40	77.5	8.1	417	10	US-09-815-242-10592	Sequence 10592, A
41	77.5	8.1	435	9	US-09-895-913A-146	Sequence 146, App
42	77.5	8.1	445	9	US-10-117-417-4	Sequence 4, Appli
43	77.5	8.1	445	9	US-10-117-417-15	Sequence 15, Appli
44	77.5	8.1	513	10	US-09-741-148A-2	Sequence 2, Appli
45	77.5	8.1	533	10	US-09-815-242-11612	Sequence 11612, A

ALIGNMENTS

RESULT 1
US-10-068-080-3
; Sequence 3, Application US/10068080
; Patent No. US20020115591A1
; GENERAL INFORMATION:
; APPLICANT: Fritz, Christian
; APPLICANT: Youngman, Philip
; TITLE OF INVENTION: USE OF YNES, ESSENTIAL BACTERIAL GENES AND POLYPEPTIDES
; FILE REFERENCE: 07334/090001
; CURRENT APPLICATION NUMBER: US/10/068,080
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/070,116
; PRIOR FILING DATE: 1997-12-31
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 193
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-10-068-080-3

Query Match	100.0%	Score	960	DB	12	Length	193
Best Local Similarity	100.0%	Pred. No.	2.6e-83				
Matches	193	Conservative	0	Mismatches	0	Indels	0
Qy	1	MLIALLLIAYLGISIPSGLVKLGKIDIREHSGNLGATNFRILGVKAGSVIAGD	60				
Db	1	MLIALLLIAYLGISIPSGLVKLGKIDIREHSGNLGATNFRILGVKAGSVIAGD	60				
Qy	61	ILKCTLTALPFLMHVDIHPHLLAGVAVLGHVFFIFAKFKGKAVATSGGVLLFYPALLF	120				
Db	61	ILKCTLTALPFLMHVDIHPHLLAGVAVLGHVFFIFAKFKGKAVATSGGVLLFYPALLF	120				
Qy	121	ITWVAFFIFLYLTKFVSLSSMLTGIYTVIYFVFDHTYLLIVVTLTIFVIYRHRANIK	180				
Db	121	ITWVAFFIFLYLTKFVSLSSMLTGIYTVIYFVFDHTYLLIVVTLTIFVIYRHRANIK	180				
Qy	181	RIINKTEPKVKWL	193				
Db	181	RIINKTEPKVKWL	193				

Qy	1	MLTALLIILAYLIGTSPGLIVGKLA	IDIREHSGSNGLGNATFTLGVKGS	VVIAGD 60
Db	1	MIITVILLIAYLLIGTSPGLIVGQV	FPQINLREHSGSNGTGTNFTFRI	LKKKAGMAFEVD 60
Qy	61	ILAKGTATLAPFLMHVD-THPLL	AGVAVLGHVFPFIFANKGKAVAT	SGCVLLFPYAPLL 119
Db	61	FFKGTLTATLPIIFHGGVSP	PLFGLLAVIGHTFPIFAGFKGKAVAT	SAGVIFGEPAF 120
Qy	120	FIPMAVFFFLYTLFVSLSSMLT	GTGYVI-----SYFV--	HDTVLLIIVTLLTF 170
Db	121	CLVLAIFGALYLGSMISLSSV	TASIAAVIGVLLPFLFCFILNSYD	SFLFIALILASL 180
Qy	171	VIVYHRANTKRIINKTEPKVK		192
Db	181	IIIRHKDNIARIKNKTENLYPV		202

RESULT 5

```

US-09-815-242-5305
: Sequence 5305, Application US/09815242
: Patent No. US20020061569A1
: GENERAL INFORMATION:
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlsen, Kari L.
: APPLICANT: Zyskind, Judith W.
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John D.
: APPLICANT: Carr, Grant J.
: APPLICANT: Yamamoto, Robert T.
: APPLICANT: Xu, H. Howard
: TITLE OF INVENTION: Identification of Essential Genes in
: FILE REFERENCE: ELITRA.011a
: CURRENT APPLICATION NUMBER: 60/9815,242
: CURRENT FILING DATE: 2001-03-21
: PRIORITY APPLICATION NUMBER: 60/191,078
: PRIORITY FILING DATE: 2000-03-21
: PRIORITY APPLICATION NUMBER: 60/206,848
: PRIORITY FILING DATE: 2000-05-23
: PRIORITY APPLICATION NUMBER: 60/207,727
: PRIORITY FILING DATE: 2000-05-26
: PRIORITY APPLICATION NUMBER: 60/242,578
: PRIORITY FILING DATE: 2000-10-23
: PRIORITY APPLICATION NUMBER: 60/253,625
: PRIORITY FILING DATE: 2000-11-27
: PRIORITY APPLICATION NUMBER: 60/257,931
: PRIORITY FILING DATE: 2000-12-22
: PRIORITY APPLICATION NUMBER: 60/269,308
: PRIORITY FILING DATE: 2001-02-16
: NUMBER OF SEQ ID NOS: 14110
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 5305
: LENGTH: 202
: TYPE: PRT
: ORGANISM: Staphylococcus aureus
US-09-815-242-5305

```

RESULT 7

US-09-823-246-2 ; Sequence 2, Application US/09823246
; Patent NO. US20020058789A1
; GENERAL INFORMATION:
; APPLICANT: Burnham, Martin K. R.

```
; APPLICANT: Biswas, Sanjoy
; APPLICANT: Zalacain, Magdalena
; APPLICANT: Warren, Patrick V.
; APPLICANT: Sylvester, Daniel R.
; APPLICANT: McDevitt, Damien
; TITLE OF INVENTION: ynes
; FILE REFERENCE: GM20001
; CURRENT APPLICATION NUMBER: US/09/823,246
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: US 60/193,496
; PRIOR FILING DATE: 2000-03-31
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 202
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-823-246-2

Query Match      48.48; Score 464.5; DB 10; Length 202;
Best Local Similarity 46.58; Pred. No. 8.8e-37;
Matches 94; Conservative 40; Mismatches 59; Indels 9; Caps 2;

Qy 1 MLIALIILAYLIGSLVGLVGLKAGIDIREHSGNLGATNAPRTLVGKAGSVIAGD 60
Db 1 MWIIVMLLSLVIGAFSGVIGKLFKKDIRQFGSGNTGATNFRVLGRPAGFLVFLD 60

Qy 61 ILKGTALATAPFLM--HVD-----IHPLLAGVFAVLGHVFFIPAKFKGKAVATSGV 111
Db 61 IFKGFITVFPWLQVHADGPISTFTFTNGLVGLFAILGHVYVYLVKFGGKAVATSGV 120

Qy 112 LFFVAPLFTIWAFFFLVLTKFVSLSSMLTGTIYTVIYFFVHDYLLIVTLLTIFV 171
Db 121 VLGVPILLILLATIFFLVKIFRYSLASIVAAICCVIGSLIIQDYILLVVSFLVSI 180

Qy 172 IYHRANKRIINKTEPKVKW 193
Db 181 IIRHSNIARIFRGEPEKIKW 202

RESULT 8
US-09-815-242-11713
; Sequence 11713, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13757
; LENGTH: 203
; TYPE: PRT
; ORGANISM: Salmonella typhi
US-09-815-242-13757

Query Match      26.88; Score 257.5; DB 10; Length 203;
Best Local Similarity 35.58; Pred. No. 2.4e-17;
Matches 70; Conservative 37; Mismatches 69; Indels 21; Caps 7;

Qy 6 LIIILAYLIGSLVGLVGLKAGIDIREHSGNLGATNAPRTLVGKAGSVIAGDILKGT 65
```

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11713
; LENGTH: 205
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-815-242-11713

Query Match      27.28; Score 261.5; DB 10; Length 205;
Best Local Similarity 35.08; Pred. No. 1e-17;
Matches 69; Conservative 39; Mismatches 68; Indels 21; Caps 7;

Qy 6 LIIILAYLIGSLVGLVGLKAGIDIREHSGNLGATNAPRTLVGKAGSVIAGDILKGT 65
Db 8 LVLLAYLCGSISSAILVCRLAGLPDRDSCSGNPGATNVLICGKGAVALIFDVLKGM 67

Qy 66 L-----ATATL---PFLMHVDIHPLLAGVFAVLGHVFFIPAKFKGKAVATSGVLLFVAPL 118
Db 68 LPVNGAWALGLTPFWLG-----LVAIACVGHIMPVFFHFRGCKGVATAGAI---API 118

Qy 119 ---LFTIWAFFFLVLTKFVSLSSMLTGTIYTVIYFFVHDYLLIVTLLTIFVIYRH 175
Db 119 GLDVTGYMACTWLLTILLGYSYSLGATVSALIAPFYVMFKPOY-TTPVSNLSCLILLRH 177

Qy 176 RANIKRIINKTEPKVKW 192
Db 178 HDNIQLWRROESKI-W 193

RESULT 9
US-09-815-242-13757
; Sequence 13757, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13757
; LENGTH: 203
; TYPE: PRT
; ORGANISM: Salmonella typhi
US-09-815-242-13757

Query Match      26.88; Score 257.5; DB 10; Length 203;
Best Local Similarity 35.58; Pred. No. 2.4e-17;
Matches 70; Conservative 37; Mismatches 69; Indels 21; Caps 7;

Qy 6 LIIILAYLIGSLVGLVGLKAGIDIREHSGNLGATNAPRTLVGKAGSVIAGDILKGT 65
```

[illegible]

RESULT 10
US-09-815-242-10997
; Sequence 10997, Application US/09815242
; Patent No. US20020061569A1

; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.

APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in Prokaryotes

Query Match	Score	DB	Length
Best Local Similarity	26.6%	255;	199;
Matches	34.0%;	Pred. No. 4e-17;	
71: Conservative	39;	Mismatches	63;
	Indels	36;	Gaps
	8:		

[illegible]

Db 166 VALVCCLLIYRHHDNIORLWRGOEDKV-W 193

RESULT 11

US-09-815-242-11799
; Sequence 11799, Application US/09815242
; Patent No. US20020061569A1.

GENERAL INFORMATION:
APPLICANT: Haseibeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.

APPLICANT: Yamamoto, Robert I.
 APPLICANT: Xu, H. Howard
 TITLE OF INVENTION: Identification of Essential Genes in
 TITLE OF INVENTION: Prokaryotes

Query Match	26.2%;	Score 252;	DB 10;	Length 189;
Best Local Similarity	32.8%;	Pred. No. 7.2e-17;		
Matches	59;	Conservative 42;	Mismatches 77;	Indels

[illegible]

RESULT 12
US-09-815-242-10303
; Sequence 10303, Application US/09815242
; Patent No. US20020061569A1

GENERAL INFORMATION:
APPLICANT: Hasebeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.

Qy	5	LLIILAYLIGTSPGSLVGVKLIAGIDIREHSGWLGATNAPFTPLGVKAGS-----VVI	57
Db	14	IFPTLLGLYGGIPFGYALMKIFYGMDITKIGSGATNTVLRALOSKGVSNAKOMALEV	73
Qy	58	AGCILKGTLTATLPMLHVDIHT-PLLAGVFAVLGLHGVPIFAKFKGSKAVATSGGVLLFVA	110
Db	74	TLPLFKGMEAFVLSKIFGLDYSIQOMWYATLSTLCVCHYSDEI NENGCKGCVSTINGSVI	130

OY 117 PLAFITWAVFFIFLYLTKFVLSUSSL-TGIYTVIYSF--FVH--DTYLLI-----VV 164
 ! : : : ! ! : ! ! : ! ! : ! : ! : ! : ! : ! : ! : ! : ! : ! :
 DB 134 PIESLGLTWFFGVKVKISSLASLGVGTATLVFFPYMHIPDSVNLKEVGQTQPM 193
 ! : ! : ! : ! : ! : ! : ! : ! : ! : ! : ! : ! : ! : ! : ! :
 OY 165 TLATFVIYRHRANIKRIINKTEPKV 190
 ! : ! : ! : ! : ! : ! : ! : ! : ! : ! : ! : ! : ! : ! : ! :
 DB 194 VLTFIFTLIKHAGNIFNLTKGKKV 219
 ! : ! : ! : ! : ! : ! : ! : ! : ! : ! : ! : ! : ! : ! : ! :

```

: APPLICANT: Yamamoto, Robert T.
: APPLICANT: Xu, H. Howard
: TITLE OF INVENTION: Identification of Essential Genes in
: TITLE OF INVENTION: Prokaryotes
: FILE REFERENCE: ELITRA.011A
: CURRENT APPLICATION NUMBER: US/09/815,242
: CURRENT FILING DATE: 2001-03-21
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/242,578
: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: NUMBER OF SEQ ID NOS: 14110
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 11453.
: LENGTH: 262

```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 24, 2002, 22:13:55 ; Search time 10196.6 Seconds
(without alignments)
1661.118 Million cell updates/sec

Title: US-10-068-080-4
Perfect score: 582
Sequence: 1 atgtaattgctttattgat.....ctaaagtaaaatggttataa 582

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rtd.*
- 36: em_htg_nam.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
C 1	582	100.0	6595	1	AF024713	AF024713 Bacillus
C 2	582	100.0	26170	1	BC170DEGR	273234 B.subtilis
C 3	582	100.0	233780	1	BSUB0010	299113 Bacillus su
C 4	176	30.2	195269	6	AX417035	AX417035 Sequence
C 5	176	30.2	349980	6	AX417044	AX417044 Sequence
C 6	171.8	29.5	250050	1	AL591978	AL591978 Listeria
C 7	168.6	29.0	7563	1	AF084044	AF084044 Listeria
C 8	121.4	20.9	9839	1	AB010565	AB010565 Fusobacte
C 9	117	20.1	25021	1	AB028896	AB028896 Streptoco
C 10	114	19.6	771	1	SGO236899	AJ236899 Streptoco
C 11	114	19.6	11545	1	AE006539	AE006539 Streptoco
C 12	114	19.6	11573	1	AE010024	AE010024 Streptoco
C 13	114	19.6	53354	1	AE014149	AE014149 Streptoco
C 14	108.2	18.6	3014	1	AF269800	AF269800 Staphyloc
C 15	108.2	18.6	3014	6	AX145118	AX145118 Sequence
C 16	108.2	18.6	4105	1	AF269437	AF269437 Staphyloc
C 17	108.2	18.6	4105	6	AX144757	AX144757 Sequence
C 18	106.6	18.3	594	1	AX144133	AX144133 Sequence
C 19	106.6	18.3	13211	1	AE006332	AE006332 Lactococc
C 20	101.6	17.5	10828	1	AE008451	AE008451 Streptoco
C 21	101.6	17.5	248254	2	SPNEU1903	AL449925 Streptoco
C 22	100	17.2	6171	6	BD003717	BD003717 Polynucle
C 23	100	17.2	6812	1	SPPARCETP	267739 Streptococc
C 24	100	17.2	10624	1	AE007390	AE007390 Streptoco
C 25	97.8	16.8	11711	1	AE013117	AE013117 Thermoana
C 26	89	15.3	295350	1	AP004826	AP004826 Staphyloc
C 27	89	15.3	303750	1	AP003133	AP003133 Staphyloc
C 28	89	15.3	346900	1	AP003362	AP003362 Staphyloc
C 29	72.8	12.5	14651	1	AE012788	AE012788 Chlorobiu
C 30	67.2	11.5	12639	1	AE009092	AE009092 Agrobacte
C 31	67.2	11.5	14140	1	AE008058	AE008058 Agrobacte
C 32	67.2	11.5	15663	1	AE001796	AE001796 Thermotog
C 33	64.6	11.1	13750	1	AE000701	AE000701 Aquifex a
C 34	62.8	10.8	146174	1	D90910	D90910 Synechocyst
C 35	61.8	10.6	8634	1	AF281816	AF281816 Acholepla
C 36	57.2	9.8	10530	1	AE013956	AE013956 Yersinia
C 37	55.2	9.5	1657	6	AX416329	AX416329 Sequence
C 38	54.2	9.3	342650	1	AP003582	AP003582 Nostoc sp
C 39	53	9.1	12216	1	AE006206	AE006206 Pasteurel
C 40	52.4	9.0	13785	1	AE005913	AE005913 Caulobact
C 41	52	8.9	293350	1	SME591786	AL591786 Sinorhizo
C 42	48.8	8.4	7218	6	I66494	I66494 Sequence 14
C 43	48	8.2	10177	1	U32713	U32713 Haemophilus
C 44	48	8.2	11198	1	AE004138	AE004138 Vibrio ch
C 45	47.4	8.1	16662	1	AE002138	AE002138 Ureaplasma

ALIGNMENTS

RESULT 1
LOCUS AF024713/C
DEFINITION AF024713 Bacillus subtilis DNA topoisomerase IV subunits ParE (parE) and
ACCESSION AF024713
VERSION AF024713.1 GI:2558945
KEYWORDS Bacillus subtilis.
SOURCE Bacillus subtilis.
ORGANISM Bacillus subtilis.
REFERENCE 1 (bases 1 to 6595)
AUTHORS Huang W.M., Libbey J.L., van der Hoeven P. and Yu S.X.
TITLE Bipolar localization of Bacillus subtilis topoisomerase IV, an enzyme required for chromosome segregation

AF024713 6595 bp DNA linear BCT 05-OCT-1999
Bacillus subtilis DNA topoisomerase IV subunits ParE (parE) and
ParC (parC) genes, complete cds.

Proc. Natl. Acad. Sci. U.S.A. 95 (8), 4652-4657 (1998)
98208596
PUBMED 9539793
REFERENCE 2 (bases 1 to 6595)
AUTHORS Huang, W.M., Libbey, J.L., van de Hoeven, P. and Yu, S.X.
TITLE Direct Submission
JOURNAL Submitted (13-SEP-1997) Oncological Sciences, University of Utah,
Salt Lake City, UT 84132, USA
FEATURES
source
1..6595
/organism="Bacillus subtilis"
/strain="CB10, a derivative of 168"
/db_xref="taxon:1423"
/map="1800 kbp"
2073..4040
/gene="parE"
/note="required for chromosome segregation"
2073..4040
/gene="parE"
/note="a subunit of DNA topoisomerase IV"
/codon_start=1
/transl_table=11
/product="ParE"
/protein_id="AAB81614.1"
/db_xref="GI:2558946"
/translation="MARKQDFDNEADIVLEGLAEVNRKPMWIGVSTDAERLHLVY
EIVNSDVEIAGHDHIIKIHKDNLSIVODRGMPGKMGKLGKPTPEVILVLA
GKFGQGYTSGGLHGVASVNALSEMLVYTERDGFVYQRFGRKPVTSLEKI
GKTGTGTLHFQDPDMFTTTFYFETLSERLESALFKLGLKIELIDERNQREVF
YENGESFAYVLAEEKDVLSEVSVFGEHRSIEVFAFQFNDCYSENLISVFNVRQT
KDGTGSEAGATATRAFNEVARKVALLKEDTDIREGLSAILISVRIPEELQ
FEGQTKLGTSEARSADVAIVSEQLAYFLEENRDTATLLVKAIKASQAREARKAR
EARSGRKKRSEATLVSGLTKPAGSRNPKNELLYVEDSGAGSAGKQRRDFQAVLQ
LRGVINTEKAKLADIKNEEINTIHAIGGVGADFSIDIDYDKIIMTDADGA
HIQVILLTTFYRKPLTEHGKVFIALPPLYKVGSKGKKEIIIEYASDEEMGVLLK
VKGVTIORYKGLGEMNADQLWETTMNPESRTLVRVKIDDAARVRRVTTLMGDKVEP
RRKTEKNVAFGLDEESNLENLSVAEEV"
4044..6464
/gene="parC"
/note="required for chromosome segregation"
4044..6464
/gene="parC"
/note="a subunit of DNA topoisomerase IV"
/codon_start=1
/transl_table=11
/product="ParC"
/protein_id="AAB81615.1"
/db_xref="GI:2558947"
/translation="MSQPELPHDLPLLEEVIGDRFGYSKYIIQDRALPDARDGLKPVQ
RRLIYAMTDCNTFDKRNKAAKTVGNVIGNYHPHGDSVYEAMVRKSDHKNVRLI
EMHNGNSIDGPPAAMRYTEARLSPIASELLRDLIDKNTEVFVNFDDTSKEPVVLP
MFPNLLVNGSTGISAGYATDIPPHLGEIVDAVIRKIQMPCSVDELMEVIGKPDFPT
GGIIQGVGIRKAYETGKGIIRKAEIETIRGREQIVITEIPFEYKNANLVKMD
EFRIDKVEGISEYVERDTRGLRVVIELKEADAKGLNLFYKNTDLOITYNFNVMA
IHNRRMLKSLPSILDVIGHQKEVTVNRSVLEQAKDRHHIIVEGLMKALSILDEVI
ATIRSSDKRAKNLIIAKYEETEPQAEIVSLQLYRLTNDITALEAEELCKKIE
ELESISNDKLLAVITNSLAKKAKYADTRSVIEEIKINLEWVASEDYVIT
VTKGILKTSORFASNGSGDKMDRMLHOFEMNTDVLILLTKNGSYIYCPVH
QLPDIRKMDGCHFSNLITIDRETIVKAIPEKFDPSAYLLFTFKNGMKVKTILTHY
KAQRYSKALVALNKGEDELIDVHTNGESQIFMATHGLYGLWFGEDVNVGARAAG
VKNILKEDDFVSGEIIQQSDSIVLFTORGAVKRMISLEFEKTSRAKRGVYMLRELK
KNPHRVVAFACGLLEORLMAETEKGRLEQTKELQTNDRYSNGSFFPDEESGKVT
VWRLHTEQ"
BASE COUNT 2096 a 1261 c 1543 g 1695 t
ORIGIN
Query Match 100.0%; Score 582; DB 1; Length 6595;
Best Local Similarity 100.0%; Pred. No. 2.le-137;
Matches 582; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGTTAATTCCTTTATTTGGCTACTCTTATAGGAGCAGCATTCCTATCGGCTTA 60
|||||
DB 1097 ATGTTAATTCCTTTATTTGGCTACTCTTATAGGAGCAGCATTCCTATCGGCTTA 1038

QY 61 ATTGTGGCAGCTTGCARAGGAATTGATATTCGGGAGCAGGAAGCGCACTTAGGC 120
|||||
DB 1037 ATTGTGGCAGCTTGCARAGGAATTGATATTCGGGAGCAGGAAGCGCACTTAGGC 978
|||||
QY 121 GCTACCAATGATTCCTGCTACATTCGGGTGTAAGAGCTGGTTCGTCATAGCCGAGAT 180
|||||
DB 977 GCTACCAATGATTCCTGCTACATTCGGGTGTAAGAGCTGGTTCGTCATAGCCGAGAT 918
|||||
QY 181 ATTTTGAAGGACATGCGCACTGCATTCCTTTCTCATGCATGTTGATATTCACCCG 240
|||||
DB 917 ATTTTGAAGGACATGCGCACTGCATTCCTTTCTCATGCATGTTGATATTCACCCG 858
|||||
QY 241 CTTCCTGCAGGAGTCCTTTGCGGTTTATAGCCACAGCTGTTTCCCATCTTCGCAAAATTTAAA 300
|||||
DB 857 CTTCCTGCAGGAGTCCTTTGCGGTTTATAGCCACAGCTGTTTCCCATCTTCGCAAAATTTAAA 798
|||||
QY 301 GCGGTAAGCCGTGGCGACATCAGGAGCGTGTTCCTCATCTTCGCAAAATTTAAA 360
|||||
DB 797 GCGGTAAGCCGTGGCGACATCAGGAGCGTGTTCCTCATCTTCGCAAAATTTAAA 738
|||||
QY 361 ATCAGATGCTGGCGGATTCCTCATCTTTTATACCTGACTAAATTTGTTCTCTCTCA 420
|||||
DB 737 ATCAGATGCTGGCGGATTCCTCATCTTTTATACCTGACTAAATTTGTTCTCTCTCA 678
|||||
QY 421 TCGATGTTAAAGGATCTATCTCTTATATATAGTTTCTTCCATGATACGATATTTA 480
|||||
DB 677 TCGATGTTAAAGGATCTATCTCTTATATATAGTTTCTTCCATGATACGATATTTA 618
|||||
QY 481 TTGATGTCGTTACCT 540
|||||
DB 617 TTGATGTCGTTACCT 558
|||||
QY 541 CGAATTATCAATAAAGACAGCTTAAGCTAAAGTAAATGTTTAA 582
|||||
DB 557 CGAATTATCAATAAAGACAGCTTAAGCTAAAGTAAATGTTTAA 516
|||||

RESULT 2
LOCUS BC10DEGR/c 26170 bp DNA linear BCT 24-JUN-1998
DEFINITION B.subtilis DNA (26.2 kb fragment; 170 degree region).
ACCESSION Z73234
VERSION Z73234.1 GI:1405443
KEYWORDS aconitase; alst gene; ccdA gene; endo-1,4-beta glucanase;
endo-1,4-beta xylanase; grIA gene; grIB gene; tlpA gene;
transketolase; yneA gene; yneB gene; yneE gene; yneF gene; yneI
gene; yneJ gene; yneK gene; yneL gene; yneP gene; yneQ gene; yneR
gene; yneS gene; yneT gene; yneC gene; yneF gene; yneG gene; yneH gene.
SOURCE
ORGANISM Bacillus subtilis
REFERENCE 1 (bases 1 to 26170)
AUTHORS Rose, M. and Entian, K.D.
TITLE New genes in the 170 degrees region of the Bacillus subtilis genome
encode DNA gyrase subunits, a thioredoxin, a xylanase and an amino
acid transporter
JOURNAL Microbiology 142 (Pt 11), 3097-3101 (1996)
MEDLINE 9124194
PUBMED 8969507
REFERENCE 2 (bases 1 to 26170)
AUTHORS Rose, M.
TITLE Direct Submission
JOURNAL Submitted (19-JUN-1996) Rose M., Johann Wolfgang
Goethe-Universitaet Frankfurt, Institut fuer Mikrobiologie,
Marie-Curie-Str. 9, Geb. N250 Frankfurt/M. GERMANY D-60439
COMMENT Overlapping sequence: X87845 (bases 3016-6779).
FEATURES
source
1..26170
/organism="Bacillus subtilis"
/strain="168"
/db_xref="taxon:1423"
82..399
gene


```

/db_xref="SWISS-PROT:045058"
/translation="MWRNASMHNSKRNDANDFDSMDWLRQFFEDFAWYDETLPIDIL
YTSQYIIEADLTFLQPTQVTVTLSGCGEFTILTVAKSSQTFEKQMLPFYFNDKNQIQ
ECENQIILTAVANKETEDGSSFSLQPLS"
gene
8356..11085
/genes="citB"
CDS
8356..11085
/genes="citB"
/codon_start=1
/transl_table=1
/product="aconitase"
/protein_id="CAA97599.1"
/db_xref="GI:1405454"
/db_xref="SWISS-PROT:P09339"
/translation="MANEQKTAAKDVQFQARKFTTTNGKTYHYVLSKALSDSGIGKYSK
LPYSIKVLSELYRQVGDGPFVKKHEVENLAKWGTAELKDKIDVFPKSRVLIODFTGVP
AVVDLAKRKAAVAGSDPKNPEIPVDLVIDHSVOVDKAGTDEALAVNMDLEFFERN
AERYKFLWAKAFKNFYQAVPATGIVHOVNLEFLASVYHAAIEDEGLVTPDITLVGT
DSHTIMINGELVGMWGVGIEAEGMLGQPSYFVPEVIGAKLVGKLPNGTATDIALNL
KYTVLNRKGVGVKVFYFEGPGIAPLADRIATIANMPEYGATCGCFPPVDEEALNVL
RLTRGRDPEHIDVVEAYCRSNGLFYTPDAEDPQFTDVVEIDLSQIEANLSGPKRPQDILI
Query Match 100.0%; Score 582; DB 1; Length 26170;
Best Local Similarity 100.0%; Pred. No. 2,1e-137; Indels 0; Gaps 0;
Matches 582; Conservative 0; Mismatches 0;
QY 1 ATGTTAATTCGTTATTGATTATTTTGGCCCTACTTGTATAGCAGCATTCATCGGCTTA 60
DB 14177 ATGTTAATTCGTTATTGATTATTTTGGCCCTACTTGTATAGCAGCATTCATCGGCTTA 14118
QY 61 ATTGTGGCAAGCTTGCCAAAGGAATTGATATTGCGGAGCACGGAAGCGCAACTTAGCG 120
DB 14117 ATTGTGGCAAGCTTGCCAAAGGAATTGATATTGCGGAGCACGGAAGCGCAACTTAGCG 14058
QY 121 GCTACCAATCATTCCTACATTCGGGTGTAAGAGCTGGTTCGGTCGTATGATATTCACCCG 180
DB 14057 GCTACCAATCATTCCTACATTCGGGTGTAAGAGCTGGTTCGGTCGTATGATATTCACCCG 13998
QY 181 ATTTGAAAGGGACACTGGCAACTGCATTCGCCTTTTCTCATGCATGTTGATATTCACCCG 240
DB 13997 ATTTGAAAGGGACACTGGCAACTGCATTCGCCTTTTCTCATGCATGTTGATATTCACCCG 13938
QY 241 CTTCCTTCAGAGAGTCCTTTGCGGTTTATAGGCCACGTGTTTCCCACTCTTCGCGCAAAATTTAAA 300
DB 13937 CTTCCTTCAGAGAGTCCTTTGCGGTTTATAGGCCACGTGTTTCCCACTCTTCGCGCAAAATTTAAA 13878
QY 301 GCGGTTAAAGCCGTGGCGACATCAGAGAGCGGTTTGGCTATTTTACGCAACCCGCTTATTT 360
DB 13877 GCGGTTAAAGCCGTGGCGACATCAGAGAGCGGTTTGGCTATTTTACGCAACCCGCTTATTT 13818
QY 361 ATCAGATGGTTGCGGTTATCTTCATCTTTTATACATTGACTAAATTTGTTCTCTCTCA 420
DB 13817 ATCAGATGGTTGCGGTTATCTTCATCTTTTATACATTGACTAAATTTGTTCTCTCTCA 13758
QY 421 TCGATGTTAAACAGGATCTATACCTGTTATATATAGTTTCTTTGTCCATGATACGATATTTA 480
DB 13757 TCGATGTTAAACAGGATCTATACCTGTTATATATAGTTTCTTTGTCCATGATACGATATTTA 13698
QY 481 TTGATTCGTTACCTTCCTACCTATTTTGTGATATACAGACACCGAGCGAACAATTTAAA 540
DB 13697 TTGATTCGTTACCTTCCTACCTATTTTGTGATATACAGACACCGAGCGAACAATTTAAA 13638
QY 541 CGAATTATCAATAAAGAACACCTAAAAGTAAATGGTTATAA 582
DB 13637 CGAATTATCAATAAAGAACACCTAAAAGTAAATGGTTATAA 13596

```

RESULT 3
 BSUB0010/c
 LOCUS
 Bacillus subtilis complete genome (section 10 of 21): from 1781201
 to 2014980.
 accession 299113 AL009126

CDS
/gene="pkSB"
882..1559

/gene="pkSB"
/function="involved in polyketide synthesis"

/codon_start=1
/transl_table=11

/protein_id="CAB13592.1"
/db_xref="GI:2634092"

/db_xref="SPTREMBL:O34769"

/translation="MNLTYKVHPKTRYOGWNYCYIIEDIVSRSAIVDPSPHLSKI
TTLTSEAEALKAVALTHSHVHNVLDPITKMFNAQVYMSKKEIDYQFRCNLLISL
DDHDTISIGNRAQCLITDHTAGGNCYLFSFIFTGDTVFTEGCGICEDDGSSAEEM
FDSIQRIKISVSPHVRYPVPGHSPGSHSIKDYOHNIYFQIDKKEYFYFKPTRNQ
KGIFDFK"

gene
1932..2798

/gene="pkSC"
1932..2798

/gene="pkSC"
/function="involved in polyketide synthesis"

/codon_start=1
/transl_table=11

/protein_id="CAB13593.1"
/db_xref="GI:2634093"

/db_xref="SPTREMBL:O34825"

/translation="MITYVPGGSGQSGKGMGLDFEFKELTDQADEILGYSIKRLCL
ENPYSNLKTOFTOPALYVYNALSYLKKIRDEEVKDFVAGHSIGEYNALFAAEADF
ETGLQVRRKGLMSLLSNGMAAIVGLNEEQVAKALKYHLHDVDIANVNPYQIYI
SGKDEKXAASLFEFTMETVTVPLNVSGPHSRYMNRKAKEEFEBFLHAFYSPPSI
PVLSNYAKPYTFEYFKKQTLADQINHSVKWTDISYILMKKGHMEFEVPGVGNVLTGLI
HRIKDAEAMPR"

gene
3302..4192

/gene="pkSD"
3302..4192

/gene="pkSD"
/function="involved in polyketide synthesis"

/codon_start=1
/transl_table=11

/protein_id="CAB13594.1"
/db_xref="GI:2634094"

/db_xref="SPTREMBL:O34877"

/translation="MNEPLVFMFSGGQSYHMGKELFKENTVFRSMLEMDIAAARR
IGTSIVEEIHFGKRVSDPFDLSILFHPAIFMIEYSLYKLEDRGIYPDYVLGSVWEN
SCDKGMALAILDKPOLNDHPOLFNSSELISINDSHFVSGEDHIRKIMEDLKEKO
ILCOLLPVSVAHSSLIDPAESAYAEFLRSKSPQKPSIPIVSSLSLTCGLHVMDENFV
NAVKKPMFREAIRYLESQHTCKEFDLPGSGTLAAVFKQLIPDGSADRCSSIIPTPHQ
ELKLNMFVEYFTRPERKFR"

gene
4189..6141

/gene="pkSE"
4189..6141

/gene="pkSE"
/function="involved in polyketide synthesis"

/codon_start=1
/transl_table=11

/protein_id="CAB13595.1"
/db_xref="GI:2634095"

/db_xref="SPTREMBL:O34787"

/translation="MITYVPGGSGQSGKGMGQGLFEQYQHLTDQAOQILGYSTIEKLT
EKSLDWNHTEYTOPALYVYNALSYLKRVEETGRKPDFAGHSLGETNMAAAGADF
ETGLRLVRRGELMGRITGGMAAVIGLSKEQVTAVLEEHLRYDIDVANENTPQOIVI
SGPKTEKARAVFENTKDKVLFPLNVSGPHSRYMNEAKQVFKQYIDSFQAPLAI
PVI SNVYAEPIYIPKKNPALSALKEQRNVQAGITAEISLGSAAEFKQDYHLYAYLAGG
HRIKNEAEPIYIPKKNPALSALKEQRNVQAGITAEISLGSAAEFKQDYHLYAYLAGG
MYGIIASKEHNVKLSRAGMGFFGTGLSKEVEDAIHAIQGLGKGQAYGINLVNMM
KHTESEKMLDILLRNOVSIVEASFLVTPVLVRYRAGVKRKNQGDVICSNRLIAK
ISRPEAESKLSPAPENMLQKILGENKITMNEALLRCIPMADDICVDEADSGGHTGG
VAYSLNPAMTSLDEMMKKYQKRIYRGAAGGIGTPEAAMAAFMILGADPILFGSINO
CTVEAATSDKVKDLQOMNVQDTAYAPAGDMFESGSKVQVLYKGVFFPARANKLYEY
ORVGSIRELDKMLAEKRYFRSV"

gene
6780..8030

/gene="pkSF"
6780..8030

/gene="pkSF"
/function="involved in polyketide synthesis"

/codon_start=1
/transl_table=11
/protein_id="CAB13596.1"
/db_xref="GI:2634096"

/db_xref="SWISS-PROT:P40804"

/translation="MTKCNLPYVVTVGVGTASIGQCKEDFASLLSISRHAFOVMKRS
GROKDSRFIGAETIASLSYDRLSKMLRKASFSSRAALVTLTEAMPEAEIDDDASRI
GLVVGSGNQORENFVEYRQDSGFISPAYGLSFMDSLCLGICTDOFGITGLAVTV
GGVASSQOLAVIHAIOVLSDGEVDTCLALGALMDLSYMECEALRALGANGTKYADE
PENACRPFQNRDGFYIGSCGALVIERKETALRGLKPYAALSGWSIKLDDGNRNDP
SLGEIHLVOKALERALLPEDIDYINPHGTGSFIDGEIEELKALRACRLSHAYINATK
SITGHLSAAGIVEIISVLQMKKSALHPSRNLDHPIDDSFHWVNEKSIYSIRIKNALS
LSMGFGCMTAVCIQNTIEKGGES"

terminator
7946..7974

/gene="pkSF"
8255..9292

/gene="pkSG"
8255..9292

/gene="pkSG"
/function="involved in polyketide synthesis"

/codon_start=1
/transl_table=11

/protein_id="CAB13597.1"
/db_xref="GI:2634097"

/db_xref="SWISS-PROT:P40830"

/translation="MLITCSEGIDFGKSLSTYIHEVYGLNRCRLPEYKQACYSCTA
GFOYANVFIQSOTSPKAKALVIASDISRFLIAGGDALEDSDNSYAEPSAGAGAVAVLY
GNEPYFDIPDANGYGYEVMOTCRPIPDSEAGSDLSLSYLDLCCCEOTFLLEYORV
PGANYQDTFYIAYHTPFGMGVAKHRTMMRKVAKVTGSIETDFLTRVPLGLNYCOR
VINGMAAGALLALASTIDQRFDTPKRIGCFYSGSCSEFYSGITTPQOERQRTFG
IEKHLDRYQLSMEYEYELLFKGSMVRFGTRNNVKLDFEMIPGIMQSTQEKPRFLFEEI
SEFHRKYRWS"

gene
9280..10059

/gene="pkSH"
9280..10059

/gene="pkSH"
/function="involved in polyketide synthesis"

/codon_start=1
/transl_table=11

/protein_id="CAB13598.1"
/db_xref="GI:2634098"

/db_xref="SWISS-PROT:P40805"

/translation="MDLVYOTIKVRFOASVCYITFHRPEANNITINDTLIEECLQVLN
OCTSVTVVVLGELVEVFCFGADFOEIOEMKRGKQASSOEPLDYDLAMKLOTGPVY
TISHVGRKNAGSLGFVSATDIAIQOTASFSLSSELLFGLYPACVLPFLIRIRGROKA
HYMTLKPISVOEASEWGLDADFADSDVLLRLRLRLRLKLLKGLIAHYKQPMSSLD
HOVSRAKALTATANODMFSDPQNMGIIRYVETGQFPWEDQ"

gene
10099..10848

/gene="pkSI"
10099..10848

Query Match 100.0%; Score 582; DB 1; Length 233780;
Best Local Similarity 100.0%; Pred. No. 2e-137;
Matches 582; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCTAATTCCTTATTGATTATTTTGGCCTACTTGATAGGCAGCATTCATCTGGCCTTA 60

Db 150556 ATGTTAATTCCTTATTGATTATTTTGGCCTACTTGATAGGCAGCATTCATCTGGCCTTA 150497

QY 61 ATTGTGGGCAAGCTTGCCAAAGGAATTTGATATTCGGGAGCAGCAGGCAAGCTTAGGC 120

Db 150496 ATTGTGGGCAAGCTTGCCAAAGGAATTTGATATTCGGGAGCAGCAGGCAAGCTTAGGC 150437

QY 121 GCTACCAATGCATTCCTGCATACATTTGGGTGTAAAGCTGTTTGGTGGTCGTCATAGCCGGAGAT 180

Db 150436 GCTACCAATGCATTCCTGCATACATTTGGGTGTAAAGCTGTTTGGTGGTCGTCATAGCCGGAGAT 150377

QY 181 ATTTTGAAGGACACTGGCAACTGCATTCGCTTTTCTCATCATGTTGATATTCACCCG 240

Db 150376 ATTTTGAAGGACACTGGCAACTGCATTCGCTTTTCTCATCATGTTGATATTCACCCG 150317

QY 241 CTTCTTCAGGAGTCTTTGGGTTTTAGGCGACCGTGTTCCTCATCTTCGCAAAATTTAAA 300

|||||

QY 303 CGGTAAAGCCGTCGACATCAGGAGCGTTTCTGCTATTTTACGACCCCTGTTATTTAT 362
 II IIIII II II II II II II II II II IIIII II II IIII
 Db 116845 TGGGAAGCAGTCTGCTACTCTGCTGTTATCTCTGCTATGACCACTTTGTTTGT 116786
 QY 363 CACGATGGTTCGGGTATCTTCATCTTTTATATCTGACATAAATTTCTCTCTCATC 422
 I I I IIIII II II II II II II II II II IIIII II II IIII
 Db 116785 AGCAGCACTAGTCTGTTCTTATTAACATTAATAATCAGCAAAATATGTCGCTTATGTC 116726
 QY 423 GATGTTAACAGGATCTATCTGTTATATATAGTTCTTCTTCTGTCATCATGATTTATT 482
 IIII II II II II II II II II II IIII II II IIII
 Db 116725 AATGATTGGAGCACTTCGACGATTAATTAATCTCTTTTCATGGAGACTGGATTTAAT 116666
 QY 483 GATTGCTCTACCTGCTGCTACATTTTGTGATATACAGACACGACGCAATTAAGC 542
 IIII IIIII II II IIIII II II IIIII II II IIIII II II IIII
 Db 116665 TATCTCTGCTGCTGCTGCTCTCTTGTATTTGGCGCCACGACCAATATACCGG 116606
 QY 543 AATTATCAATAAACAAGCAACCTAAAGTAAATGTTTATAA 582
 IIII II IIIII IIIII IIIII IIIII IIIII IIIII IIIII IIIII
 Db 116605 CATTGGAACGGCGAAGAACCAAAATAAATGATGTAA 116566
 RESULT 6
 AL591978/c 250050 bp DNA linear BCT 06-JUN-2002
 LOCUS Listeria monocytogenes strain EGD, complete genome, segment 6/12.
 DEFINITION AL591978 AL591824
 ACCESSION AL591978.1 GI:16410540
 VERSION AL591978.1
 KEYWORDS Listeria monocytogenes.
 SOURCE Listeria monocytogenes
 ORGANISM Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
 REFERENCE 1
 AUTHORS Glaser, P., Frangeul, L., Buchrieser, C., Rusniok, C., Amend, A.,
 Baquero, F., Berche, P., Bloeker, H., Brandt, P., Chakraborty, T.,
 Charbit, A., Chetoui, F., Couve, E., de Daruvar, A., Dehoux, P.,
 Domann, E., Dominguez-Bernal, G., Duchaud, E., Durant, L.,
 Dussurget, O., Entian, K. D., Fsihi, H., Portillo, F. G., Garrido, P.,
 Gautier, L., Goebel, W., Gomez-Lopez, N., Hain, T., Hauf, J.,
 Jackson, D., Jones, L. M., Kaerst, U., Kretz, J., Kuhn, M., Kunst, F.,
 Kurapat, G., Madueno, E., Maitournam, A., Vicente, J. M., Ng, E.,
 Nedjari, H., Nordsiek, G., Novella, S., de Pablo, B., Perez-Diaz, J. C.,
 Purcell, R., Rammel, B., Rose, M., Schlueter, T., Simoes, N.,
 Tierrez, A., Vazquez-Boland, J. A., Voss, H., Wehland, J. and Cossart, P.
 Comparative genomics of *Listeria* species
 Science 294 (5543), 849-852 (2001)
 TITLE
 JOURNAL
 MEDLINE 21537279
 PUBMED 11679669
 REFERENCE 2 (bases 1 to 250050)
 AUTHORS Glaser, P., Frangeul, L. and Rusniok, C.
 TITLE Direct Submission
 JOURNAL Submitted (06-JUN-2001) Glaser, P., Institut Pasteur, Genomique des
 Microorganismes Pathogenes, 25 rue du Docteur Roux, 75724 Paris
 Cedex 15, FRANCE
 COMMENT E-mail: pglaser@pasteur.fr
 Phone: +33 1 45 68 89 96, Fax: +33 (0)1 45 68 87 86.
 FEATURES
 source
 1. 250050
 /organism="Listeria monocytogenes"
 /strain="EGD-e"
 /db_xref="taxon:1639"
 210..753
 /gene="lmo1125"
 210..215
 /gene="lmo1125"
 220..753
 /gene="lmo1125"
 /codon_start=1
 /transl_table=11
 /protein_id="CAC99203.1"
 /db_xref="GI:16410541"
 /db_xref="SPTREMBL:Q8Y724"
 /translation="MLPSYLVKSTAKELSLNDFLWNIIVTINKEETPEILYCGTLKEI
 EGDQITGEAYEIPKIIIAKSTLSGKEILLYDGNVGYDSMFCDFEEATQNELKK

YPINLSNRLSIGIDYSEKEEVEFDENGHVVLIDDRHIPWEQVKTGDFDFLEIT
 ATDENDSLILTEELA"
 join(764..769,778..1230)
 /gene="lmo1126"
 764..769
 /gene="lmo1126"
 778..1230
 /note="similar to E. coli Yjap protein"
 /codon_start=1
 /transl_table=11
 /protein_id="CAC99204.1"
 /db_xref="GI:16410542"
 /db_xref="SPTREMBL:Q8Y723"
 /translation="MLINKRAENODYOMILAIWEKSVIATHDFLAAREDFYKEQIP
 LFDNVLELFWFAGEIVGSGTSERELDMFLNPIATNGGSGQSLIWLIENTKRNL
 VDVEQENATRYLKHGFWVSSRDLGNGPNPYILHLQLKERVNGI"
 1208..1213
 /gene="lmo1127"
 1208..1213
 /gene="lmo1127"
 1220..1692
 /gene="lmo1127"
 1220..1675
 /gene="lmo1127"
 /codon_start=1
 /transl_table=11
 /protein_id="CAC99205.1"
 /db_xref="GI:16410543"
 /db_xref="SPTREMBL:Q8Y722"
 /translation="MEYELLONELGELEIOLBELONKEVRGMVKTAKDGTITRRL
 AKTKAGYFVAFWEKDEGKNCPEDENDACDFLCIVIDGLNGLFVFPROCLIDK
 GLTSEASGKMAARYPTWCQDLNKTAKKTONAQOQVYFANYSIRKEDH"
 1660..1692
 /gene="lmo1127"
 complement(1689..2711)
 /gene="lmo1128"
 complement(1689..2711)
 /gene="lmo1128"
 /codon_start=1
 /transl_table=11
 /protein_id="CAC99206.1"
 /db_xref="GI:16410544"
 /db_xref="SPTREMBL:Q8Y721"
 /translation="MAKNGVDITKRWKKTITMSALIVLVAIFLIGNVYEMNETRVA
 IPTGGKLSAVTTPKHKPKGLIIVFVGDAQOETONGYKPLMERFAKQGVISVSW
 DKLGVGKSSGNLNSMDRANEVNIEMMKVYKPDSTAKIGLWGAQAGVYVPKAM
 NANNEIDFSLAAPAINRWROGEYNTCWAOKDTGKTKKVKQAQNFLLDSALISKNK
 TYESYKONGKKEMTPGRYNFIKNNODATLDSKVKTKIYLALAEKDKNVDSETK
 DVTQIKKENLETQIPNVSHQINPAINSELNINIVGLMVPKLYFLVDNAYLDICE
 QVVSQK"
 complement(2666..3280)
 /gene="lmo1129"
 complement(2666..3280)
 /gene="lmo1129"
 /note="similar to unknown proteins"
 /codon_start=1
 /transl_table=11
 /protein_id="CAC99207.1"
 /db_xref="GI:16410545"
 /db_xref="SPTREMBL:Q8Y720"
 /translation="WFLSKSIDSEKDSIVERVASLVFAGFKSKTKTLFKETEARD
 ITYAFCHYITKGENLLTAIQDELVCGLFVTSKSDSYOKLYQTLRKFLPSOCLKL
 IFLLSLSHKPVQNERIDFTVSPNFRGOGIGTKLSHCLETFHDEKITLYVAKNS
 GAYKLYRNLFQVIOKENSTLMGLVGTIKERKMEWILQNVGKK"
 complement(join(3277..3996,4004..4009))
 /gene="lmo1130"
 complement(3277..3996)
 /gene="lmo1130"
 /note="similar to transcription regulators"
 /codon_start=1
 /transl_table=11
 /protein_id="CAC99208.1"

```
/db_xref="GI:16410546"
/db_xref="SPTREMBL:O8Y7Y9"
/translation="MLTSLTGEIAQLFOISKYKIRHYIDEGILVPRKRNPNNGYFFEE
IHSQEMRLNEITFFAKDDRYEKKVPNOIHDNTINYSOAAKHNLPLDLEPEYIFSK
ODGVICLKARKSDTTPAGGAYACKSFIAKDENTYETQIQLFLEELKATPTTSMILY
ENIYPSLAYPDWVYTMVEVPL"
complement(3288..3293)
/gene="lmo1129"
RBS
complement(3288..3293)
/gene="lmo1129"
RBS
complement(4004..4009)
/gene="lmo1130"
join(4233..4237,4245..5960)
/gene="lmo1131"
4233..4237
/gene="lmo1131"
4245..5960
/gene="lmo1131"
/notes="similar to ABC transporters, ATP-binding proteins"
/codon_start=1
/transl_table=11
/protein_id="CAC99209.1"
/db_xref="GI:16410547"
/db_xref="SPTREMBL:O8Y7Y8"
/translation="MIDKRLFOLVERKSLVLLILFRVLSLGLMIGLWLVFAQLTHYL
EGONVDMWLGTVLVGLVKAILTKLAEKQIYOASAEFLRSMRRVMEKAFRLGHNE
GOLPASTLTOLAVDGEIOLEIYYSRFLPOLFYCLIASLMIFGSLGVFAMQPAIVLLIC
MPMPVIMAVMKIAKRLISGYSYDYNLTGKPHENLSGLSILKAYDQYKKEEIVS
DAERFRTKMSLSMQNSITIMIIISYSGALGIGMSLIMFTNGTISMTGMLVLLL
SAEFFIPRMLQSLFVHMNGISACSLKFAYLEKQOVGSVELAKPLEKIEVHHLT
TYEGEKAKALQDYSVHFHNGISFALVSGSGKSTFVRLNQLPGYQGEIVWVPL
ASLSEGAIRKQAVDYNHGVLYANSIRENLLIGNPAANDSDLNWLVPOVSLADFVRKL
PEOLNEEDGNSLGGORQLLARALLARQAEVYVFEITSGVDLESKILLVILQ
ELAKKIVLEISHRLYNVLGADQVLFVDFAGKLLVESPRLQESYFKNYFLEEARL
LKGA"
5948..5953
/gene="lmo1132"
5948..5953
/gene="lmo1132"
5962..7608
/gene="lmo1132"
5962..7608
/gene="lmo1132"
/notes="similar to ABC transporters, ATP-binding proteins"
/codon_start=1
/transl_table=11
/protein_id="CAC99210.1"
/db_xref="GI:16410548"
/db_xref="SPTREMBL:O8Y7Y7"
/translation="MSEWTIGWLLKFKPLRGKMLAILLIGIISNLVIMISLIGTY
GIATVILQQLNPYKVLVFMVACGVVRGLARYLEQYLNHDIARLLAIIRERIPATLR
KLGARLSGKSGDLVAAITDVEALEVFHATISPVFIAGLTITATVGLASVDVGL
AIIILVQIMVGVVPLMISYKRKKIGTAYOQAFVGLNOMVENIASLODIFQFKLE
ERUKLETQEKLNQYKKRIQSGEQLIGELVGLTITAILVGLSLOLPETVLI
GTVLSUSFGSGLNALGALTITFASGRKLYALTEERKPVILFVGLSDFESAEI
DKVSFSDHQPLNELSLALPKGLIGESGKSTLVKLMRYWDPDQVKLNN
NELPKITESSLHRLGVEQRTFLFDLTGNIRLGGKDATLEEVKEAARAAIDTWI
ETLPEGYETIIGSQSRNLSGQRQIGRLARFLHDAPLLLLDEPTSNLDYINQAILN
TLRSEIQKTVLISHRATTLDAEQLFENGALKRAAK"
complement(join(7604..7637,7645..7953,7961..7966))
/gene="lmo1133"
complement(7604..7637)
/gene="lmo1133"
7615..7637
/gene="lmo1132"
7615..7637
/gene="lmo1132"
complement(7645..7953)
/gene="lmo1133"
/notes="similar to B. subtilis yjcs protein"
/codon_start=1
/transl_table=11
```

```
/protein_id="CAC99211.1"
/db_xref="GI:16410549"
/db_xref="SPTREMBL:O8Y7Y6"
/translation="MYLKENTNYLYCTAVIQTITGKYSYEQVLEHLEALRAKTVKEPGC
Query Match 29.5%; Score 171.8; DB 1; Length 250050;
Best Local Similarity 57.8%; Pred. No. 3.5e-33;
Matches 330; Conservative 0; Mismatches 232; Indels 9; Gaps 1;

QY 21 TATTTTGGCTACTTGTATAGCAGCATTCCATCTGCTTAATTTGTCGGCAAGCTTGCAA 80
Db 147579 TTTATTTAGCCCTAUGTATGTTGGTTCGATACCATCTGGTTTATGGATTGGCAAAATTTTCTA 147520
QY 81 AGGAATTTGATATTTCGGGACGACGACGCACTTAGCGCTTACCACTTACCATTCCTCCGTAC 140
Db 147519 CAAJAAAAGATATTCGCGAGTTCGGTAGCGGTAACTTAGGGCGGACAAATTCCTTCCGAGT 147460
QY 141 ATTGGGTGTAAGCTGGTTCGGTCGTCATAGCCGGAGATATTTTGAAGGACACATGGC 200
Db 147459 GCTTGGTATCAAGCTCGTAGCATTTGTACAGTAATGATATTTTAAAGGAACCTGTTGC 147400
QY 201 AACTGCATTGCTTTTCTCATGCA-----TGTGTGATATTCACCCGCTTCTTGCAGG 251
Db 147399 AACGTTGCTTCCATCTCTTTCAATTTGAATGTAGATCATCATTTCTGGTTGCTAACTGG 147340
QY 252 AGCTTTTTCGGGTTTATAGCCACGCTGTTTCCCATCTTCCGCAAAATTTTAAAGCGGTAAAGC 311
Db 147339 AGCATTTTGCATTATATCGGCATAGTTTCCCACTTTTCCGAGGATTCAGAGGTGGGAAAGC 147280
QY 312 CGTCGGACATCAGGAGCGGCTTTTGTATTTTACGACCCCTGTATTTATTCACGATGGT 371
Db 147279 AGTAGCTACTTCTCGCGAGGATTCATCTCGCTATGCGCCATTTACTCTTTGTAGCCGACT 147220
QY 372 TGCGGTATTTCTCATCTTTTATATGACTTAAATTTGTTCTCTCTCATCATGTTTAAAC 431
Db 147219 TGTAGTCTTCTTAGTTACACTAAACTTAGTAATATGTATCTCTCACTCAATGATGGTGG 147160
QY 432 AGGCATATATCTGTTATATATAGTTTCTTTGTCATGATACATGATGATTTATTTATGTTGCT 491
Db 147159 AGCACTTGTGCTAATTAATTTCTTTTCTATGCGGAGCTGGATTTTAAATTTGTTCTCGT 147100
QY 492 TACCTGCTCACTATTTTGTGATATACACACACCGGAGCGAATTAAGCAATTAATCA 551
Db 147099 TGCTTGTATCGGCTCTTTGTAATTTGCGCTACCGAGCAAAATATCACTCAATACGTAA 147040
QY 552 TAAACAGACCTAAAGTAAATGGTTATA 582
Db 147039 CGGAGAGACCAAAATAAATGGATGTA 147009

RESULT 7
AF084044/c
LOCUS
DEFINITION
AF084044
Listeria monocytogenes putative LacX gene, partial cds; putative
topoisomerase IV subunit A (parC), and putative LuxS genes,
complete cds; and unknown gene.
ACCESSION
AF084044
VERSION
AF084044.1 GI:21328241
KEYWORDS
Listeria monocytogenes.
SOURCE
Listeria monocytogenes
ORGANISM
Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
REFERENCE
1 (bases 1 to 7563)
AUTHORS
Lampidis, R., Kostrewa, D. and Hof, H.
TITLE
Molecular characterization of the genes encoding DNA gyrase and
topoisomerase IV of Listeria monocytogenes
J. Antimicrob. Chemother. 49 (6), 917-924 (2002)
JOURNAL
MEDLINE
22035523
PUBMED
12039883
REFERENCE
2 (bases 1 to 7563)
AUTHORS
Lampidis, R.
TITLE
Direct Submission
```

JOURNAL Submitted (13-AUG-1998) Institut fuer Medizinische Mikrobiologie und Hygiene, Universitaet Heidelberg, Fakultae fuer Klinische Medizin Mannheim, Klinikum Mannheim gGmbH, Theodor-Kutzer-Ufer 1-3, Mannheim 68167, Germany

FEATURES Location/Qualifiers

source 1. .7563

organism="Listeria monocytogenes"

strain="EGD"

/db_xref="taxon:1639"

<1..314

/note="ORFA"

/codon_start=3

/transl_table=11

/product="putative Lax"

/protein_id="AAM48490.1"

/db_xref="GI:21328244"

/translation="GAARYLPLNYDLFKNDALIFEALKOKEMTIKSDKTPHFVKVSFP

EPFVGWTKAKTFLPCIEPWYGIADGAGESVELRDKAGIEHLEPEAVFASEYIT

G"

CDS complement(356..952)

/note="ORFB"

/codon_start=1

/transl_table=11

/product="putative YneS"

/protein_id="AAM48493.1"

/db_xref="GI:21328247"

/translation="MTINILLSLAYVIGSIPSLGWICKFKSTKKDIRFEGSGNLGAT

ISFRVIGRAGSIVVMVLKGTAVTLFFQNLNVDHFWLLTGAFAIGHSFFLFA

GRFGKAVATAGVILAYAPLVAFLVFLTKLSYVLSLSSMIGALAAALIIISLFM

GWILIVVACALFVIMRHRANITIRNGEPPKIKWM"

1110..1517

/note="ORFC"

/codon_start=1

/transl_table=11

/product="putative YneF"

/protein_id="AAM48491.1"

/db_xref="GI:21328245"

/translation="MONPTKEEIQHLLKKAIVAVVGLSDKPDRTSYRYSQIMQOAGY

KIIPVNPDAEILGKTVKHLTDIEHVDIVNVRSEFLPEIAEAFDIDADVFAQ

IGIANQAEFLTEKGYPTVMDICIKVATQEMD"

1698..3665

/gene="parE"

1698..3665

/gene="parE"

/codon_start=1

/transl_table=11

/product="DNA topoisomerase IV subunit B"

/protein_id="AAM4848.1"

/db_xref="GI:21328242"

/translation="MNRNEYNDSDIQVLEGLAVRKRPAMYIGSTDVRLHLVVEIV

DNSVDTLAGFGKKIVYTLHEDGSVSDGEGMPVGMHKTGKSTVEVILTVLHAGGK

FOEGGKTSGLHGVGSSVNVNALSFWLVTINRDGATYQOKFKDGRKPDGTLKKIGK

SKATGTTIRFKDPALFPPTSYNFTLSERLRESAFLKGLMLIELDERVGMAEAFH

EGKNFVEYINEGKDLVHPVAFEGENATIEVENAFQNDGYSNLSIFVNNVTRG

AGSHESGMAAKTRIFNDYARRVNLKDKNDGSDIREGLSALVLSIRVPKLIQFE

GOTKEKKGTOEAROAVDAVVAEHLAYFLAENPETSLLVKKAVKAREAREARKAREE

TRNGKKKRSFLLSKGLPAQSRPNKKNELLYVEGDSAGGSAKQKQRRRFOALPLR

TRNKTERAKLDILKNEISITIIHTVGAGVTEFDEDCNDKVVIMTDADGARI

QVLLTFYFVWRPLVAGKVFIALPPLYKVSRGSKKEVIEYAMTDELDLSAQIKG

KYMIQRYKGLGEMNADQLWETTPDPTLIRVRVDDSSARAERRVATLMDGKVEPR

QWKEHVEFSMEDSQNILENNENNVBEAKDI"

3662..6121

/gene="parC"

3662..6121

/gene="parC"

/codon_start=1

/transl_table=11

/product="DNA topoisomerase IV subunit A"

/protein_id="AAM48489.1"

/db_xref="GI:21328243"

/translation="MSNPEQHIODLALEEVMGDRFGRYSKYIIQERALPDVDRGLKPV

QRRLFLAMVEGNTAEKFRKSAKTVGNVIGNYHPHGDSSVYEAVMYRMSQDMKVRNL

IEMHGNGSVGDGPPAARMRYTEARLSPISAEILLRDIKETVDIFPNFDDTSSEPTVLP

ARFPNLVNGSTGISAGYATDIPPHNLTEIEAVIKRLDNPLCTTDDIMKIVKGPDPF

TGGIIOGIDGIRKAYOTGKRVVRSKTEIEDIRGRKOITITHEPYEVNKANLVKRM

DELREKTEGISEVRDETRTGLRIANVELKDKANAEGLVNLVFNKTDLOVSYNEMV

AINKRPELMGIIIPMLDAYIEHQKEIITKRSEYDIRKAROHILEGLIKALSILDEV

IKLIRGSKDRKALNLQTKYDFSEKQAEIVSLQIRLUTNTDIIHELHSEAKSLAEQI

SVLEKILGDAEELIAVLKEELAEIKKRYKTRARRTEVOAEIKIKIDTEVLVANEIV

SVTKGYVYKTSORSYAAASNGAELAMKEADHAIQIKMNSLDSLLFTSKGNFIYRPV

HELPIRWKNLGDHVSHLASDLISAGEEISATIAOFTKEKFLVTKNGMTQKSAIT

NYKPOYSKSMMAIKLKGDELLSVHLIDGTEDIPLATKNGYGLVYSITEIPESGART

AGVKAINLKQDDIYIGGIVLMPNEQKHILLATQRGSLKQMKASEPEPISRAKRGLLML

RELKSNPHRFIGITLADNNDHLFIETNTDQIIVEIDVANLRITDRYSNGSFVLDETMEG

EPTSIWLAIEPIKDTSAKEE"

6204..6671

/note="ORFD"

/codon_start=1

/transl_table=11

/product="putative LuxS"

/protein_id="AAM48492.1"

/db_xref="GI:21328246"

/translation="MAEKMNVESFNLDHTKVKAPFVRLAGTKVGVHGDIEYKYDVRFK

QPKHEMEMPALHLEHLMAELARNHTDKLDVISPWCOTGYFVSFINHSDYDDALEI

IATLTDLVLPATEVPACNEVQCNAASHRLEGAKALAEFLAKRSEKRVNPOE"

7001..57563

/note="ORFE"

/codon_start=1

/transl_table=11

/product="unknown"

/protein_id="AAM48494.1"

/db_xref="GI:21328248"

/translation="NMSIKSNIMKICVSMVVLPSLTSLPSFAAEETALKESODVV

NIPDQLKKNLYLNOAENADITEAOMFNITLNSGIKDLTGIEYLNKNTTILSIN

NINASYEPIOTLSLEKLVINGENVNADIIPDLSVLSNLTFLDLSRTNDDTIFSKIS

NIPNLKLDISNNKIGITKISLENNLPS"

BASE COUNT 2638 a 1298 c 1575 g 2052 t

ORIGIN

Query Match 29.0%; Score 168.6; DB 1; Length 7563;

Best Local Similarity 57.4%; Pred. No. 2.5e-32;

Matches 328; Conservative 0; Mismatches 234; Indels 9; Gaps 1;

QY 21 TATTTGGGCTACTTGATAGGACGACATTCATCTGGCTTAATTTGTGGGCAAGCTTCCCAA 80

DB 926 TTTATTAGCCTATGTGATTTGGTTCGATACCATCTGCTTTATGATGGAATTTTCTAC 867

QY 81 AGCAATTGATTTTCGGGAGGACGAGCGCAACTTAGCGCTTACCAATGCAATTCGCTAC 140

DB 866 AAAAAAGATATTTCCGCGAGTTCGGTAGCGGTAACTTAGGGCGGACAAATTTCTTCCGAGT 807

QY 141 ATTGGGTGTAAGAGCTGGTTCGGTCTGCTATAGCGGAGATATTTTGAAGGGGACACTGGC 200

DB 806 GCTTGGTATCAAGCGTGGTAGCATTTGTACAGTAATGGTCATTTTAAAGGAACACTGTGC 747

QY 201 AACTGCATTTGCCCTTTTCTCATGCA-----TGTGTATATTTCACCCCGCTTCTTCAGG 251

DB 746 AACGTTGCTTCCATCTTCTTTCAGTTGAATGTAGATCATCATTTCTGTTGTTGTAACG 687

QY 252 AGTCTTTGGGTTTATAGGCGACGCTTTCCCATCTTCGCCAAATTTAAAGGCGGTAAAGC 311

DB 686 AGCAATTGCTATTTATCCGCGATAGTTTCCACACTTTTTCGAGGAGTTCAGAGTGGGAAGC 627

QY 312 CGTGGCGACATCAGGAGGCGTTTGTATTTTACGACCCCTCTTATTTATCATCAGTGT 371

DB 626 AGTAGTACTTCTGCGGAGTCAATCTCGCCTATGCGCCATTAATCTTTGTAGCCGCACT 567

QY 372 TCGGCTATTCTTCATCTTTTATATGACTTAAATTTGTTCTCTCTCATCGATGTTAAC 431

DB 566 TGTAGTCTTCTTAGTTACACTAAAACTTAGTAAATATGATCTCTCAGTTCAATGATTGG 507

QY 432 AGCGATCTATACGTTTATATATAGTTCTTTTGTCCATGATACGATATTTATGATTCGT 491

DB 506 AGCACTTGCTGCATTAATTTCTCTTTTTCATGGGGGACTGGATTTTAAATTTCTTCGT 447

QY 492 TACCTGCTCACTATTTTCTGATATACACACACCGGCGCAACATTAACGAATTAACAA 551

QY	61	ATTGTGGCGAAGCTTGCCAAAGGAATTTGATATATTCGGGAGCAGCGAAGCGGCAACTTAGGC	120
Db	7424	TGGATAGGAAAAGCTTTTAAAGGTGTTTCATGTAAGAGATTAATGAAGATPAAAAATAGTGGT	7365
QY	121	GCTACCAATGCATTCGCGTACATTCGGGTGTAAGAGTCGTCGGTGCATAGCCGGAGAT	180
Db	7364	GCTACAACCTCATATAGATTTTATAGGACAAATTTGGGTGATGCTGTTTAAATATGAT	7305
QY	181	ATTTTGAAGAGCACATGGCAACTGCATTCGCTTTTCTCATGCAT---GTTGATATATTCAC	237
Db	7304	GTCPTAAAGAGTTTTATACCTCTTTATATATCCAGTAAATTTAATTTAGTATATAATCAT	7245
QY	238	CCGCTCTTCGACAGAGTCCTTTCGGGTTTATAGGCACAGTGTTCCTCCATCTTCGCCAAATTT	297
Db	7244	TTATGTTATTTTATGTTAGTTAGCAATTTTAGCTCATACATATTTTCTTGCTTATATCTTT	7185
QY	298	AAAGCGGTAAAGCGTGGCGACATCAGGAGCGTTTTGCTATATTTTACGCACCCCTGTTA	357
Db	7184	AAAGTGGAAAGGTGTTGCAACTAGTTTATAGAGTTTCTTATTTTAAATACCTGTTTATA	7125
QY	358	TTTATCAGATGGTTCGGGTATCTTCATCTCTTTTATATCTTGACTAAATTTGTTCTCTC	417
Db	7124	ACTTTAATATATTAGCAATTTTATTTTACTTGCTTATTTTACTAAATATGTTTCATG	7065
QY	418	TCATCGATGTTAACAGGAGTCATACACTGTTATATATAGTGTTCCTTGTGTCAT---GATACG	474
Db	7064	GCATCTATAACAGCAGCTTTTTTATGTCCTATTTTACTTTTTTTCACATCAATAAGATTC	7005
QY	475	TATTTATTTGATGTCGTTTACCTGCTCACTATTTTGTGATATACAGACCGAGCGAAC	534
Db	7004	TATTTATTTGCTTTATCAGTTTATATAGCAGTTTTTGTGTTATATATAGACATAAGACA	6945
QY	535	ATTAAGCAATTTATCAATAAAACAGACCTAAAGTAAATGGTTA	581
Db	6944	ATTTCAAGATTACTTAGTGGACAGAAATAAATTCAAATTTTAATA	6898
RESULT 9			
AB028896/c			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
REFERENCES			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
FEATURES			
SOURCE			
genome			

```

CDS
/gene="cpsIas"
<1..375
/gene="cpsIas"
/function="putative histidine protein kinase"
/codon_start=1
/transl_table=11
/product="CpsIas"
/protein_id="BAB62877.1"
/db_xref="GI:15186710"
/translation="DHYDTWKFKELKESNHPVLIAFGERHWDRLTSGKSLAECLQLTD
LDEVKSTIQLOREFKSVRNPLAHLIKPFEQELRYTQFSQAFLDQIIFLAKVIG
VEYTVNFHVDYVKNKLIKILE"
complement(400..1323)
/gene="cpsIar"
complement(400..1323)
/gene="cpsIar"
/function="LysR-type transcriptional regulator"
/codon_start=1
/transl_table=11
/product="CpsIar"
/protein_id="BAB62878.1"
/db_xref="GI:15186711"
/translation="MRIQQLQYVIVIKIVETGSMNEAAKOLVITOPSLSNVNRNLETEMG
IQFIRNPKGITLTKDMEFLSVARQILEOTALLEERYKGDNTSRELFVSSSHAYEV
VNAFVALNGTDYQELFELREHRTWELIDVKNFRSEIGULFNLSTNRDVLKLFDD
NSLIATLFTTTHIFVSKSNPLANRKLKSKMKDLEDPYLSIDOGSLHNTIYKIKKI
QIHPKTSIVSDRATFLNLMIGLDGYTVATGILNSKLNGDEIVAIPLDVDDVIDIVI
RHKANLSKMGQKFDYLLLEVSFNRKNKQS"
complement(1447..1452)
/10_signal
evidence=experimental
1453..1457
/35_signal
evidence=experimental
complement(1470..1475)
/35_signal
evidence=experimental
1478..1482
/10_signal
evidence=experimental
complement(1481..1486)
/10_signal
evidence=experimental
complement(1507..1512)
/35_signal
evidence=experimental
1512..2969
/gene="cpsIaa"
1512..2969
/gene="cpsIaa"
/function="regulator"
/codon_start=1
/transl_table=11
/product="CpsIaa"
/protein_id="BAA82275.1"
/db_xref="GI:5381178"
/translation="MSNHSRRQKKSHSTPLRVINLFLVIFILLSVYSLFLMYRRHF
LAFRLNAVYGVTVLVLILASLFLCIKNKARIEFTIILVLASFVATTLDFKFKSTIDL
TNLNKNTASYSEIEMSVIVPKDSKITNEAVSKLAAPVKNDTSNITDLIEHKSEKGI
SITPKQTSQODANRIRKNGDSQAMVNNAYVSLIELSTPDFKSOIKITVYKIKKI
NRKTNKHEGVNIYISGIDTFSISVSRSDVINIMVNTNHHKVLTLTTPRDAVYK
IPDGGNQYDKTHAGYLVGVSMTKLENLYDINLDYARINFSSFLKLIDLGGVTY
YNDQAFYTKHGNDFPVQVTLNSEQALGVFRERYSLQGGDNRGRNQRVIAIINK
LASSQVYTKLNSITSLQTSVQTNMTIDNINDLNQLSTQCRFTVESQALTGHGVTG
ELPSYAMPGAQLYMSIDQSSLSNAKSKIKNTWEE"
2975..3706
/gene="cpsIab"
2975..3706
/gene="cpsIab"
/codon_start=1
/transl_table=11
/product="CpsIab"
/protein_id="BAA82276.1"
/db_xref="GI:5381179"
/translation="MIDIHSHVFDVDDGPKTLEESLSLIEESYRQGVRIIVSTSHRR
KGMFETEDIFKHSIVKHEAKRFEHLQILYGGELYTSDMLEKLLKLOIPTLNNT
KFALIEFSMOTSKNDIHTALSNNVLMGITPVVAHIERYNALENOKERVKEIINNVCYT
QINSNHLKOKLFNDKHKRKRYRLEENLVHFVASDMHNLDRPFPFLAEAYKIC
RDFGKERANQLFIENAOISILKNHYI"

gene
3715..4407
/gene="cpsIac"
3715..4407
/gene="cpsIac"
/function="chain length determination (putative)"
/codon_start=1
/transl_table=11
/product="CpsIac"
/protein_id="BAA82277.1"
/db_xref="GI:5381180"
/translation="MNKIANTEVEINFLKLLWKMKKFLITFVAIAFATAGLPYSILF
IVTPOYISSPRIYVINTPNTPNNSITADQLQAGSFLANDYKEIITSYDVLKYSSEKL
NYPSSQLQKITVLSIKDTRVISVEDANPNMSQKLANYSREARVSKIKAVTQVEDI
TTLEKGNLKPAPSSPNIKKNVLGIVGAGLSTIVLIMGILDDRNVNTEEDIEKALGL
TSLGIVPDLNKL"
4418..14922
/misc_feature
4418..5107
/10_signal
evidence=experimental
4418..5107
/10_signal
evidence=experimental
5243..5248
/35_signal
evidence=experimental
5265..5270
/35_signal
evidence=experimental
5334..6509
/gene="cpsIae"
5334..6509
/gene="cpsIae"
/function="glucosyltransferase"
/codon_start=1
/transl_table=11
/product="CpsIae"
/protein_id="BAA82279.1"
/db_xref="GI:5381182"
/translation="MWLKYSFYVIFISSLFFIFPKNSFTTTRLSFFSFTAMNSILLYL
LNSFLATYRKYSYAKFSRDTKVVLTITNKDSLKMTFRNKYDHNHYIAVCILDSSEKDY
DLKHNLSRIINKDALTSCLTVDQAFINPIELFKYQIOIDIINDIEAMGVIVN
VEALSFDNIGEKRIQTFEGYSVITYSMKFYKSHLIAKRLDITGAIIGLIGIVAI
FLVPOIRKDGGAIFSONRVGRNGRIFRFYKFSMRVDAEQIKKDLLVHMQTGLMFK
LDDPRITKIGTIRKTSIDELPQYVNLKGDMSLVGTRPTVDEYKYNSTOKRRLS
PKGIVGLWQISGRNNITDFDEIVKLDVQYINENWSIWSDIKILLTLKVVLLGTGAK"
6533..6982
/gene="cpsIaf"
6533..6982
/gene="cpsIaf"
/codon_start=1
/transl_table=11
/product="CpsIaf"
/protein_id="BAA82280.1"
/db_xref="GI:5381183"
/translation="MKICLVGSSGGHLAHLNLLKPFCKNKDRFWVTDFKEDARSILRE
EIVYHGFPTNRNVKNLVKNTILAFKVLKRDPDVLISCAAVAVPFFYIGKLGKCT
VYIEVFDRIKPTLTGKLVYPVTDKFLVQWEEKKVPKRAINLGGIF"
6982..7524
/gene="cpsIag"
6982..7524
/gene="cpsIag"
/function="galactosyltransferase"
/codon_start=1
/transl_table=11

```

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 771)
AUTHORS Vriesema,A.J.M.
TITLE Direct Submission
JOURNAL Submitted (08-FEB-1999) Vriesema A.J.M., Medical Microbiology,
Academic Medical Center, University of Amsterdam, Room L1-160,
Meibergdreef 15, 1105 AZ Amsterdam, NETHERLANDS

FEATURES
source Location/Qualifiers
1..771
 /organism="Streptococcus gordonii"
 /strain="CH1"
 /db_xref="taxon:1302"
gene complement(1..176)
 /gene="parE"
CDS complement(<1..104)
 /gene="parE"
 /codon_start=1
 /transl_table=11
product="topoisomerase IV"
protein_id="CAB40548.1"
db_xref="GI:4584072"
translation="MAKKEININYNDDAIOVLEGLDAVRKRPGMYIGS"
RBS complement(111..116)
 /gene="parE"
-10_signal complement(150..155)
 /gene="parE"
-35_signal complement(171..176)
 /gene="parE"
-35_signal 202..207
note="ORF2"
-10_signal 224..229
note="ORF2"
CDS 267..>771
note="ORF2"
codon_start=1
transl_table=11
product="hypothetical protein"
protein_id="CAB40549.1"
db_xref="GI:4584073"
translation="MINTILGLLAVLGSIPTGLWIGOIFKKNLREYSGNLTGTTNN
TRFLIKTAGTDTFADFLDKGTLATLLPLEHLINGISPMIFGLIINVLFHPFPFAEK
GGKAVATAGVXMGFSPLFFSLIIIFIIVTLYGSMISLASIVVAGFAIILSVLIPLP
GIIIPSYD"

BASE COUNT 215 a 147 c 121 g 287 t 1 others

ORIGIN

Query Match 19.6%; Score 114; DB 1; Length 771;
Best Local Similarity 55.1%; Pred. No. 2e-18;
Matches 264; Conservative 0; Mismatches 211; Indels 4; Gaps 2;

OY 20 TTATTTCGGCTACTTGATAGGCAGCACTTCCACTCGCTTAATTGTGGCAAGCTTGCCA 79
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 289 TAATTTGGCATATTTACTGGGTTCAATTCACACTGGTTATGGATTGGACAGATTTC 348
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 80 AAGAAATTGATATTCGGGAGCACGGAACGCGCAACTTAGCGCTACCAATGCATTCCGTA 139
|| | | | | ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 349 TTAATAAGAATTTACGAGATACGGTAGTGSAAATACGGGCACAACACACTTCCGTA 408
|| | | | | ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 140 CATTTGGGTGTAAGACTGTGGTTCGTCATAGCCGGAGATATTTTGAAGGGACACTGG 199
|| | | | | ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 409 TTTTAGGAAGACAGACTGGAACATGATCTTTTGGCAATCGACTTTCTAANAAGGAAC 468
|| | | | | ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 200 CAACTGCATTGCCTTTTCTCATGCATCT---TGATATTCACCCGCTTCTTCACGAGTCT 256
|| | | | | ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 469 CAACCTACTCCACTTTTTTCTCCATATCAATGATATCAACCATGATTTTGTCTAA 528
|| | | | | ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 257 TTGCGGTTTTAGGCCAGCTGTTTCCCCTATCTTCGCCAAATTTTAAAGGGCGGTAAGCCGTGG 316
|| | | | | ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 529 TAGCTGTCTTGCCACACCTTCCCAATTTTGTCTGAATTTAAGGAGGCAAGCCGTTG 588
|| | | | | ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 317 CGACATCAGAGGGGTTTTGCTATTTTACGCACCCCCCTGTTATTTATCACGATGTTGCGG 376
|| | | | | ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 589 CTACAAGTCGAGGTGTTGTCNTGGATTTTCTCCCTCTCTCTTTCCCTATCTGATTTATTA 648
|| | | | | ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 377 TATTCTTCATCTTTTATCTACTGACTAAATTTGTTTCTCTCTCATCATGCTTACACGGA 436
 Db 649 TTTTATTTGTAACCCCTTACTTGGGAAGTAGTATGATTTCTTTGGCTAGCATTTGTTAGCTG 708
 QY 437 TCTACTGTTTATATATAGT-TTCTTTGTCCATGATACGATATTATTGATTGTCGTTAC 494
 Db 709 GCTTTGGCATATTTCCTGCTCTTATATTCTCTTTATTAGGAATTATCTCTCCTAGTTAC 767

 RESULT 11
 AE006539/c
 LOCUS 11545 bp DNA linear BCT 01-JUN-2001
 DEFINITION Streptococcus pyogenes M1 GAS strain SF370, section 68 of 167 of
 the complete genome.
 ACCESSION AE006539 AE004092
 VERSION AE006539.1 GI:13622059
 KEYWORDS
 SOURCE
 ORGANISM Streptococcus pyogenes M1 GAS.
 Streptococcus pyogenes M1 GAS
 Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 Streptococcus.
 REFERENCE 1 (bases 1 to 11545)
 Ferretti,J.J., McShan,W.M., Adjic,D., Savic,D., Savic,G., Lyon,K.,
 Primeaux,C., Sezate,S.S., Surorov,A.N., Kenton,S., Lai,H., Lin,S.,
 Qian,Y., Jia,H.G., Najjar,F.Z., Ren,Q., Zhu,H., Song,L., White,J.,
 Yuan,X., Clifton,S.W., Roe,B.A. and McLaughlin,R.E.
 Complete genome sequence of an M1 strain of Streptococcus pyogenes
 Proc. Natl. Acad. Sci. U.S.A. 98 (8), 4658-4663 (2001)
 21192684
 11296296
 REFERENCE 2 (bases 1 to 11545)
 Ferretti,J.J., McShan,W.M., Adjic,D., Savic,D., Savic,G., Lyon,K.,
 Primeaux,C., Sezate,S.S., Surorov,A.N., Kenton,S., Lai,H., Lin,S.,
 Qian,Y., Jia,H.G., Najjar,F.Z., Ren,Q., Zhu,H., Song,L., White,J.,
 Yuan,X., Clifton,S.W., Roe,B.A. and McLaughlin,R.E.
 Direct Submission
 Submitted (10-APR-2001) Department of Microbiology and Immunology,
 University of Oklahoma Health Sciences Center, 940 SL Young Blvd,
 Oklahoma City, OK 73104, USA
 LOCATION/Qualifiers
 1. 11545
 /organism="Streptococcus pyogenes M1 GAS"
 /strain="SF370"
 /serotype="M1"
 /db_xref="taxon:160490"
 362..1075
 /gene="deod2"
 /note="Best Blastp hit = sp|Q56037|DEOD_STRTR PURINE
 NUCLEOSIDE PHOSPHORYLASE (INOSINE PHOSPHORYLASE) (PNP)
 >gi|1276873|gb|AAC44007.1| (U40830) Deod [Streptococcus
 thermophilus] >gi|1588804|prf|12209356a deod gene
 [Streptococcus thermophilus]"
 /codon_start=1
 /transl_table=11
 /product="putative purine nucleoside phosphorylase"
 /protein_id="AAK33813.1"
 /db_xref="GI:13622060"
 /translation="MSIHSAKKGIADKILLPCDPLRAKFAIENFLEDAVCNEVRN
 MFGYTGTYGHRVSVMTGCMGHSISYARELIVDYGVTLIRVGAGALDPVHVRE
 LVLAARAAANNIIRNDFPEFDFGLDKAYHIAEMGVTTVHGVNLSDDVY
 TNNPRMALGKLGVAIEMEAALYLLAAQHHVKAALGIMTISDNLDPTEDTTAEER
 QTFTDMKVGLETLIAND"
 1068..1856
 /gene="SPY0895"
 1068..1856
 /gene="SPY0895"
 /note="Best Blastp hit = gb|AAB81305.1| (L47121) histidine
 protein kinase [Carnobacterium piscicola]"
 /codon_start=1

/transl_table=11
 /product="histidine protein kinase"
 /protein_id="AAK33814.1"
 /db_xref="GI:13622061"
 /translation="MTNQTLDLDDVAYNHAFAKALPNIPKTYLXLLLEMKERR
 EUNLAFIAEHAENRTIEDQYHCSLNQSLDEQYHFOAMLESDHVKNGAIIIDFVRS
 VSPILYFLRLITSEIPNFQYVIFDTKNDQYDTWHFOAMLESDHVKNGAIIIDFVRS
 NYTKSLADMLTILSLPOELKDLVFLRHFEKAVRNPLAHLIKPFDEELHRTHTFS
 QAFLENIITLFTSGVIYRREPFYDDMNAIINKELSLWRQSIY"
 complement(1935..2840)
 /gene="cpsy"
 /note="SPY0898"
 complement(1935..2840)
 /gene="cpsy"
 /note="Best Blastp hit = emb|CAB36977.1| (Y17221) Cpsr
 protein [Streptococcus agalactiae]"
 /codon_start=1
 /transl_table=11
 /product="putative transcriptional regulator"
 /protein_id="AAK33815.1"
 /db_xref="GI:13622062"
 /translation="MRIQULHYIIKIVEGSMNEAAKQLFITOPSLSNVAKVLEMEHG
 ITIFNPKGITLTKDGVFELSYARQIIETQSLLEDYKHNHTGRELFSVSSQAFY
 VNAFVLLKRTDMTRYELFRTRETWIIEIDVKNFRSEIGVLFINDYNDVLTGKLFDD
 NHTASPLKFAOPHIFVSKSNPLATKSLSMODLRDPVLSYDOGHNSFYFSEENMS
 QHPHNSIVVSDRAFTFLNLMIGLDGYTVASGILNSNLNGDQIVAIPLDVPDEIDIVI
 KHEKANLSRMGERFIYLLLEEVTFD"
 3062..3598
 /gene="SPY0899"
 3062..3598
 /gene="SPY0899"
 /note="Best Blastp hit = dbj|BA882291.1| (AB028896) ORF1
 [Streptococcus agalactiae]"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AAK33816.1"
 /db_xref="GI:13622063"
 /translation="MNVTEQNSHOILQKLVSIHYLTFRDELKLVERTPTSLGGEF
 PAHVQSEGLDIAAIDTDMOORLIESTFYEESEAFKLMNKTLDIVDMWIGVDHLI
 DICQSKVEFQIIIGDKRIRVFGVLSDFSLKVSALSKEARIPDLYEHIQVRNLEE
 EAFIKYQSPKVPTEGE"
 3875..4567
 /gene="pyrF"
 /note="SPY0900"
 3875..4567
 /gene="pyrF"
 /EC_number="4.1.1.23"
 /note="Best Blastp hit = sp|Q92HA7|DCOP_STRPN OROTIDINE
 5'-PHOSPHATE DECARBOXYLASE (OMP DECARBOXYLASE) (OMPDCASE)
 >gi|4009483|gb|AAC95452.1| (AF068902)
 orotidine-5'-decarboxylase PyrF [Streptococcus
 pneumoniae]"
 /codon_start=1
 /transl_table=11
 /product="putative orotidine-5'-decarboxylase PyrF"
 /protein_id="AAK33817.1"
 /db_xref="GI:13622064"
 /translation="MKERPIIALDFSSFEETKAFLDLFPAAEKLYVKIGMELYAAG
 PDIVRYIKSLGHNVFLDHLDPNTVRAAMVVKELDIDMATVHAAGVEMLKAARE
 GLGQPTLIATVLTQSTSEDOMRGDQIOTSLLESVLYHVSKGAAKQALDGVACSQDEV
 EAKVATPVTGCTLTGIRPKSGNIGDQKRVWTPNQARRIGSDYIVVGRPTQAKDPV
 AYQAIKAEWAG"
 4625..5254
 /gene="pyrE"
 /note="SPY0901"
 4625..5254
 /gene="pyrE"
 /EC_number="2.4.2.10"
 /note="Best Blastp hit = gb|AAC95453.1| (AF068902) orotate
 phosphoribosyltransferase PyrE [Streptococcus pneumoniae]"
 /codon_start=1
 /transl_table=11

Db	11450	ATGAAATTTACTACTTTTATTTATACCAATTCCTTATTTACTTGGTCTATTTCACACTGGACATA	11391
Qy	61	ATTGTGGCGAAGCTTGGCAAGGAATTCATATTCGGGAGCACGGAAGCGCAACTTAGGC	120
Db	11390	TGGATTGGACAGTACTTTTACCACATCAACTTACGAGAGCATGCATCAGGAAATACTGGA	11331
Qy	121	GCTACCAATGATTCCTGATACATTCGGGTGTAAGAACTGGTTCGGTCTGCATAGCCGGAGAT	180
Db	11330	ACCACAAATACTTTTTCGGATTTTATAGTGTCAAGGACGGAACAGCTACCTTAGCTATTGAT	11271
Qy	181	ATTTTGAAAGGACATGGCAACTGCATTCGCTTTTCTCATGCATGT---TGATATTACAC	237
Db	11270	ATGTTTTAAAGGACATTTTCAATATGTGTTACCAATATTTTGGTATGACTTCAAATTTCA	11211
Qy	238	CCGCTTCTTCGAGGAGTCTTTTCGGGTTTTAGGCCACAGTGTTCCTCATCTTCGCCAAATTTT	297
Db	11210	TCCATGCTATCGGCTTTTTCGCACTTTTAGGGCATACTTTTTTCCCTATTTTGGCCAACTTT	11151
Qy	298	AAAGCGGTAAAGCGGTGGCGACATCAGGAGGCGCTTTTGCTATTTTATAGCACCCCTGTTA	357
Db	11150	AAAGTGTGTAGGCGGTAGCAACAAGTCTGGTGTATTTGCTTAGGCTTTGCTCCGTATAT	11091
Qy	358	TTTATCAGGATGGTTTCGGGTATTTCTCATCTCTTTTATCTATGACTAAATTTGTTCTCTC	417
Db	11090	CTCTTTTTTTTAGCATCTATCTTTGTTTACTTCTCTATTTATTTAGTATGATATCTTTA	11031
Qy	418	TCATCGATGTTTAAACGGGATCTATACCTGTATATATATAGTTTCTTTTGTCATGATACGTAT	477
Db	11030	GCTAGTGTGGTTTCAGTATCTGTTGGTGTGTATCTGTTTAAACATTTCTCTGCCATTCAT	10971
Qy	478	TT 479	
Db	10970	TT 10969	
RESULT	i2		
LOCUS	AE010024/c		
DEFINITION	Streptococcus pyogenes strain MGAS8232, section 72 of 173 of the complete genome.	11573 bp	DNA linear BCT 03-APR-2002
ACCESSION	AE010024		
VERSION	AE010024.1		
KEYWORDS	GI:19748120		
SOURCE	Streptococcus pyogenes MGAS8232.		
ORGANISM	Streptococcus pyogenes MGAS8232		
REFERENCE	Bacteria: Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.		
AUTHORS	1 (bases 1 to 11573) Smoot,J.C., Barbian,K.D., Van Gompel,J.J., Smoot,L.M., Chaussee,M.S., Sylva,G.L., Sturdevant,D.E., Ricklefs,S.M., Porcella,S.F., Parkins,L.D., Beres,S.B., Campbell,D.S., Smith,T.M., Zhang,Q., Kapur,V., Daly,J.A., Veasy,L.G. and Musser,J.M.		
TITLE	Genome sequence and comparative microarray analysis of serotype M18 group A Streptococcus strains associated with acute rheumatic fever outbreaks		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (7), 4668-4673 (2002)		
MEDLINE	21927593		
PUBMED	11917108		
REFERENCE	2 (bases 1 to 11573)		
AUTHORS	Smoot,J.C., Barbian,K.D., Van Gompel,J.J., Smoot,L.M., Chaussee,M.S., Sylva,G.L., Sturdevant,D.E., Ricklefs,S.M., Porcella,S.F., Parkins,L.D., Beres,S.B., Campbell,D.S., Smith,T.M., Zhang,Q., Kapur,V., Daly,J.A., Veasy,L.G. and Musser,J.M.		
TITLE	Direct Submission		
JOURNAL	Submitted (31-JAN-2002) Laboratory of Human Bacterial Pathogenesis/Rocky Mountain Laboratories/NIAID/NIH, 903 S. 4th St., Hamilton, MT 59840, USA		
FEATURES	Location/Qualifiers		
source	1. .11573		
gene	/organism="Streptococcus pyogenes MGAS8232" /strain="MGAS8232" /db_xref="taxon:186103" 362..1075		

FEATURES	hemilton, MI 39840, USA
source	Location/Qualifiers 1. .11573 /organism="Streptococcus pyogenes MGAS8232" /strain="MGAS8232" /db_xref="taxon:186103" 362..1075
gene	

```

/feature="deco"
/feature="spym18_0955"
362..1075
/feature="deco"
/feature="best blastp match gb|AAK33813.1| (AE006539)
putative purine nucleoside phosphorylase [Streptococcus
pyogenes M1 GAS]
/codon_start=1
/transl_table=11
/product="putative purine nucleoside phosphorylase"
/protein_id="AAL97596.1"
/db_xref="GI:19748121"
/transl_table="MSIHSAKSGDIADKILLPCDPLRAKFAIENFLEDAVCFNEVRN
MFGVTGYKGRHVSVMGTGMGSPISIIYARELIVDGVKTLIRVGTAGADPVDHVRE
LVAQRAATNLIIRNDFPEFDPQIADFGLLDKAYHAREMGVTHVGNVLSDDFY
TNPNRNALGKGVRAIENEAALYLAHQHVKALGIMTISDNLNDPTEDTIAEER
QTFTDMKVGLETLIAND"
1062..1856
/feature="spym18_0956"
1062..1856
/feature="spym18_0956"
/feature="best blastp match gb|AAK33814.1| (AE006539)
histidine protein kinase [Streptococcus pyogenes M1 GAS]
/codon_start=1
/transl_table=11
/product="histidine protein kinase"
/protein_id="AAL97597.1"
/db_xref="GI:19748122"
/transl_table="MPMTNOTDILDVYVYAHAFRIAKAPNIPTKVLVYLLEMLKE
RRELNLAFLAHAENRTIEDVYHCSLWLOSLEDEJIANIYILDLEVKVNGAIIDFV
RSVPIYRFLPLITSEIPNFKAYIPFTKNDQVDTWHFOAMLESDEHEVKAYLSOKO
SRNVTTSKADMLTSLPQEIKNLVLFLRHFEKAVENPLAHLIKPEDEELHRTTHF
SSQAFLENIITLAFSGVIYRPEPFYDDMNAIKKELSLWRSIV"
complement(1936..2841)
/feature="cpsy"
/feature="spym18_0957"
complement(1936..2841)
/feature="cpsy"
/feature="best blastp match gb|AAK33815.1| (AE006539)
putative transcriptional regulator [Streptococcus pyogenes
M1 GAS]
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator"
/protein_id="AAL97598.1"
/db_xref="GI:19748123"
/transl_table="MRIQQLHYIKIIVECGSMNEAAKQLEFIPQSLSNVAKDLEMENG
ITIFNRNPKGITLTKDGEFELSVARQIIQTSLEEDRYKNHNTGRELFVSSQHYAFV
VNAFVSLKRTDMTRYELFRETRETWIIDDVKNFRSEIGVLEINDYNRDVLTKFDD
NHLTASPLFRAOPHIYVSKSNPLATKSLSDMDLRDPYLSYDQGIHNSFYFSEEMMS
OMPHNKSIVVSDRATLFLMIGLDGTVASGILNSNLNGDQIVAIPLDVPDEIDVFI
KHEKANLSKGERFIEYLLEEVFD"
3063..3599
/feature="spym18_0958"
3063..3599
/feature="spym18_0958"
/feature="best blastp match gb|AAK33816.1| (AE006539)
conserved hypothetical protein [Streptococcus pyogenes M1
GAS]
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAL97599.1"
/db_xref="GI:19748124"
/transl_table="MNVTEQNSHOILOKLLVSHYITLFRDELKLVERTPSILGGFF
PAHLVQSELGDIIVAAIDTLDQOQLIESTFYAESAFKLNKMTLIDVNNIKGVYHLI
DLQSKGVFQIIGDKRIRVFGVLSDFSSLSKVSALSKEAPIPDVLYEHIQVNLLE
EAFIKYQSPKVPPTGEF"
3876..4568
/feature="pyr"
/feature="spym18_0959"
3876..4568
/feature="pyr"

```

```

/feature="best blastp match gb|AAK33817.1| (AE006539)
putative orotidine-5'-decarboxylase PyrF [Streptococcus
pyogenes M1 GAS]
/codon_start=1
/transl_table=11
/product="putative orotidine-5'-decarboxylase"
/protein_id="AAL97600.1"
/db_xref="GI:19748125"
/transl_table="MKEERPIALDFSSPEETKAFLDLPAAEKLVYKICMELYAQQ
PDIVRIKSLGHNVFLDLKHLIDPNTVRAMAVLKELDIDMATVHAAGVEMUKAARE
GLGQPTLIATVOLTSTSEDMRGDQNIQTSLSLESHYSKGAARQDLGVVCSAQEV
EAIKATVPTGFTCTLPTGIRPKGSGNDGKRVMTPNQARRIESDYIVVGRPIQAKDPV
AAQQAIAKAWAG"
4626..5255
/feature="pyr"
/feature="spym18_0960"
4626..5255
/feature="pyr"
/feature="best blastp match gb|AAK33818.1| (AE006539)
putative orotate phosphoribosyltransferase [Streptococcus
pyogenes M1 GAS]
/codon_start=1
/transl_table=11
/product="putative orotate phosphoribosyltransferase"
/protein_id="AAL97601.1"
/db_xref="GI:19748126"
/transl_table="MTLASQIATQLDIDKAVYLKPEDPFTWASGIKSPYITDNRVTL
YKPTROLDENGFEVETIRAHFPEVEIAGTATAGIPIHGAIIADKMTLPFYAIRSKPKDH
GAGNOITEGVLKQKNVIEDISTGGSVLDAASAREGADVLGVVAFIYELPKAS
ONFKEAGIKLITLSNVTETLIAVAKLQGIYITDGLHLLKKFKEDQVNNQQ"
5452..6906
/feature="amic"
/feature="spym18_0961"
5452..6906
/feature="amic"
/feature="best blastp match gb|AAK33819.1| (AE006539)
putative amidase [Streptococcus pyogenes M1 GAS]
/codon_start=1
/transl_table=11
/product="putative amidase"
/protein_id="AAL97602.1"
/db_xref="GI:19748127"
/transl_table="MTYQDATAMAIQVOTGPTPLELVTOAIYKAKKLNPLNATISE
RFEALAEAKORDFSGLPFAGVPIFLKDLGOELKHSSTSGSRLFKFYQATKTDLFVK
RLAALGFILGRSNTPEFGFNKISDSLSHGPVNLKPDNTRNAGSGSGGAALVSSGIS
ALAPASGGGSIKIPASFNGELKGSRGMPVPGPSYRSWQGSVHFALTKSVROTR
NLLYIQMSOMESPFLATLTKDSIYQSQRPLTIAPYORLSGDSVPSLDTAKALQOA
MTWAIYQSGKDIIPAWRYSOVLQKWDYTSATMASFHYETYNLLTFTTNTAPKRGELVP
DSKLMANLAQAEIIFSEEDQNLVETMFGKSLAINPYTALPNLTGQPAISLPTVETREG
LPMGIOLIAAKGREDLLLGIAEQFEANGLLKIPQ"
7024..7887
/feature="spym18_0962"
7024..7887
/feature="spym18_0962"
/feature="best blastp match gb|AAK33820.1| (AE006539)
putative ABC transporter (binding protein) [Streptococcus
pyogenes M1 GAS]
/codon_start=1
/transl_table=11
/product="putative ABC transporter (binding protein)"
/protein_id="AAL97603.1"
/db_xref="GI:19748128"
/transl_table="MKKSTLLSIGLASLLLAACAPHKSKQSSWNTIKKGVKLKATP
GTYQPTSFYNDNNELGYEVDYMKKEIKRLNIKVKFVTFDQAFVSDVSGRYDISLN
NFIDTPKRQKYNISTPYKYGVGMIVRADGSSNMKADLSMKKGKKAAGASGTYMK
VAQKQCAEMVYDNTVYLVNDVANGRTDFIPNDYPAOKLFVDYMLSONPMLNKMMS
DVQYNTEQGIVMNKKDSSLKXIDDAVIDMKKDGSLKISSETIYAGODLTKEPFGDK
KIPVIDTKVN"
7916..8566
/feature="spym18_0963"
7916..8566
/feature="spym18_0963"

```

```
/note="best blastp match gb|AAK33821.1| (AE006539)
putative ABC transporter (permease) [Streptococcus
pyogenes M1 GAS]
/codon_start=1
/transl_table=11
/product="putative ABC transporter (permease)"
/protein_id="AAL97604.1"
/db_xref="GI:19748129"
/translations="MINPLMKDSLGFVLSPLVTIGISLSPFTGLVFGLLALLGR
SROPLHLYRTYISINRGVPMIVLVFGLPYGYLEPALLCALVGLGSMVSAAYI
SEVRSIEAIDGQWAAKALGLPYALMKVKKILPOAFRIAPVPGNVIIDMWKSS
LAAMITVPDIFQNAKIIGREWDMYSYIIVAFIYWLIAFLLERYOEFLKALV"
8699..9352
/gene="ung"
/note="spym18_0964"
8699..9352
/gene="ung"
/note="best blastp match gb|AAK33822.1| (AE006539)

Query Match          19.6%; Score 114; DB 1; Length 11573;
Best Local Similarity 53.7%; Pred. No. 1.8e-18;
Matches 259; Conservative 0; Mismatches 220; Indels 3; Gaps 1;

Oy 1 ATGTTAATGCTTTATGATTTATTTGGCTTACTTGATGACGATTCCTATCGGCTTA 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11478 ATGAATTAATGCTTTATGATTTATTTGGCTTACTTGATGACGATTCCTATCGG 11419

Oy 61 ATTTGGGCAAGCTTCCCAAGGAATTCATTTCCGGAGACGGAAGCGGCAACTTAGGC 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11418 TGGATTGGACAGTACTTTTACCACATCACTTACGAGAGCATGGATCAGGAATACTGGA 11359

Oy 121 GCTACCAATGATTCCTCGTGTAAAGCTGCTTCCGTCGTATAGCCGCGGAGAT 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11358 ACCACAAATGATTTTCGGATTTAGGTGTCAAGCAGCAAGCATCTTAGCTATTGAT 11299

Oy 181 ATTTTGAAGGACACTGGCAACTGCATTTGCTTTCTCATGCAATGTTTCAATTTCCAC 237
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11298 ATGTTTAAAGGACACTTTCAATATGTTACCAATTTATTTTGGTATGACTTCAATTTCA 11239

Oy 238 CCGCTTCTTGAGGAGTCTTTGCGGTTTTTGGCCACGCTTTTCCCATCTTCGCCAAATTT 297
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11238 TCCATTGCTATCGGCTTTTTCGCGATTTTAGGCGATACATTTTTCCTATTTTTCGCCAACTTT 11179

Oy 298 AAAGCGGTAAAGCGGTGCGGACATCAGGAGCGTTTTCGTATTTTACGCAACCCCTGTTA 357
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11178 AAAGTGTGTAAGCGGTGAGCAACAGTCTGCTGTATTCGATAGGCTTTCCTCGCTTATAT 11119

Oy 358 TTTATCAGATGGTTCGGGTATTTCTCATCTTTTATATCTTGACTAAATTTGTTTCTCTC 417
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11118 CTCTTTTTTTAGCATCTATCTTTGTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 11059

Oy 418 TCATCGATGTTAACAGGATCTATCTGTTATATATAGTTTCTTGTGCTCATGATCGTAT 477
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 11058 GCTAGTGTGGTTTCAGCTATCGTTGGTGTGTATCTGTTTAAACATTTCTCGCCATCAT 10999

Oy 478 TT 479
    ||
Db 10998 TT 10997

RESULT 13
AE014149/c
LOCUS
DEFINITION Streptococcus pyogenes MGAS315, section 14 of 37 of the complete
genome.
ACCESSION AE014149 AE014074
VERSION AE014149.1 GI:21904329
KEYWORDS
SOURCE Streptococcus pyogenes MGAS315.
ORGANISM Streptococcus pyogenes MGAS315.
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE 1 (bases 1 to 53354)
```

```
AUTHORS Beres,S.B., Sylva,G.L., Barbian,K.D., Lei,B., Hoff,J.S.,
Mammarella,N.D., Liu,M.-Y., Smoot,J.C., Porcella,S.F.,
Parkins,L.D., McCormick,J.K., Leung,D.Y.M., Schlievert,P.M. and
Musser,J.M.
TITLE Genome sequence of a serotype M3 strain of group A Streptococcus:
Phage-encoded toxins, the high-virulence phenotype, and clone
emergence
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (15), 10078-10083 (2002)
REFERENCE 2 (bases 1 to 53354)
AUTHORS Beres,S.B., Sylva,G.L., Barbian,K.D., Lei,B., Hoff,J.S.,
Mammarella,N.D., Liu,M.-Y., Smoot,J.C., Porcella,S.F.,
Parkins,L.D., McCormick,J.K., Leung,D.Y.M., Schlievert,P.M. and
Musser,J.M.
TITLE Direct Submission
JOURNAL Submitted (14-JUN-2002) Laboratory of Human-Bacterial Pathogenesis,
Rocky Mountain Laboratories, NIAID, NIH, 903 South Fourth St.,
Hamilton, MT 59840, USA
FEATURES
source
1..53354
/organism="Streptococcus pyogenes MGAS315"
/strain="MGAS315"
/serotype="M3"
/db_xref="taxon:198466"
/note="group: A"
123..986
/gene="thya"
/note="Spym3_0601"
123..986
/gene="thya"
/note="best blastp hit: gb|AAK33802.1| (AE006538) putative
thymidylate synthase [Streptococcus pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="putative thymidylate synthase"
/protein_id="AAM79208.1"
/db_xref="GI:21904330"
/translations="MLTERLEMTKADIFKANIQKIINEGSLSEQARPKYKDGRTAH
SKYITGAFAYDLAKGEFFITLRPIKSAIKELFWIYQDQNSLDVLEAKYNVHW
NEVEDOTRTIGORYGAVKKHDIISKILQALAEWPNRRNVLISMDYAEFTKGLL
PCAFQIMFDVRRVGEDLYLDASLTRSDILVAHHINAMQYVALQMMIAKHFGKIKG
FFYFVNLIHYDNQFOAQELLKRPVQSPKLVLYNPDRTNFFDIKPDDDFELQNVDP
VKQLHFDLAI"
1066..1563
/gene="dvr"
/note="Spym3_0602"
1066..1563
/gene="dvr"
/note="best blastp hit: gb|AAK33803.1| (AE006538) putative
dihydrofolate reductase [Streptococcus pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="putative dihydrofolate reductase"
/protein_id="AAM79209.1"
/db_xref="GI:21904331"
/translations="MTKEIIAIWADEAGLIGVAGLPWYLPKLEBHFKKTTLHQAIL
MGVTFEGMNCRLPOROTLVMTNRDYOVDLVTMTSIEKVLWYHAQDKTLYIIGG
NKVLEAFNGYFDRIIKTVIHREFKGTYRNLDFSHFTQESQFYADAKANPDFTVT
ALXHK"
1583..1753
/gene="Spym3_0603"
1583..1753
/gene="Spym3_0603"
/note="best non-GAS blastp hit: gb|AAK75657.1| (AE007451)
hypothetical protein [Streptococcus pneumoniae TIGR4], and
gb|AAL00232.1| (AE008512) hypothetical protein
[Streptococcus pneumoniae R6]"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAM79210.1"
/db_xref="GI:21904332"
/translations="MQRSIFGVFTFLVCVICVLCATPAFKKNRYGLASFLFLNAFTNL
VNSIHAFYGTLF"
```

```

1883. 3112
/gene="clpX"
/note="Spym3_0604"
1883. 3112
/gene="clpX"
/note="best blastp hit: gb|AAK33805.1| (AE006538) putative
ATP-dependent Clp protease subunit x [Streptococcus
pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="putative ATP-dependent Clp protease subunit x"
/protein_id="AAW79211.1"
/db_xref="GI:21904333"
/translation="MAGSRNDIKVYCSFCGKSODDVKKIAGNNVFCNECVALSOE
IIRKEAEVLADTEVPKRELDVLNQYVGGDRAKRALSVAYNHYRVSFTESR
DDDDVDLQKYNLMIGPTGSKTFLAOTLAKNSLNPFAIADATSLTEAGYVGEDENI
LLKLIADYNVERAERGIIYVDEIDIAKNGENSIYTRDVSQGVQOALLKIEGTV
ASVPQGRKHPNEMIQIDTKNLFIVGGAFDGEIEIVKORLGEKIVGFQNSRKID
DNASYMCEIISDIOKEQLIPEFIRGLPVVALEOLNTSLIOILTEPRNALVQVOA
LLSDGVELAFDREALEAIANKAIERKTGAGRLRSIIETWLDIMFEIPSOEDVTKVR
ITKAAVEGSKPVLETA"
3122. 3721
/gene="Spym3_0605"
3122. 3721
/gene="Spym3_0605"
/note="best blastp hit: gb|AAK33806.1| (AE006538)
conserved hypothetical GTP-binding protein [Streptococcus
pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="putative GTP-binding protein"
/protein_id="AAW79212.1"
/db_xref="GI:21904334"
/translation="MAEEQVLNTHNASILLISAANKSHYPQDDLPETIAGLRSNVKGSS
FINILGRKNLARTSSKPGKTLQNFNIDDKLRFVDVPGYVAKVSKSRKWKRM
EYLTGRNLRAVSLVLDLRHAPSKEDIQWDFLKYDIPVIVVATKADKIPHGKWK
HESVVKALNFDKSDTFIVFSSVERIGIDDSWDAILEQV"
3869. 4615
/gene="Spym3_0606"
3869. 4615
/gene="Spym3_0606"
/note="best blastp hit: gb|AAK33807.1| (AE006538)
hypothetical protein [Streptococcus pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAW79213.1"
/db_xref="GI:21904335"
/translation="MTOVAQGRSKGKSLCHOCNALFVFIIFLILLAVLFGVRYRNS
LEGWRTSIDOKLDDFAKRLTGLHOSPLIDDSLLSSQMLITVKNNDVLSFVQV
ERDIFARLAAHYONELLKTLKENHLVVGDLSSKEROIENSMPASHELEMLDQAFE
KLASQIGGKYNQKTHLSAVVLLGKGNRIHLTIDIEEVAAGHTSFSGKLLTPNGYFD
YTRFGKLELIGDEKIIIFKALKKSVRCIT"
complement(4668. 6767)
/gene="clpL"
/note="Spym3_0607"
complement(4668. 6767)
/gene="clpL"
/note="best blastp hit: gb|AAK33808.1| (AE006538) putative
ATP-dependent Clp protease (ATP-binding subunit)
[Streptococcus pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="putative ATP-dependent Clp protease
(ATP-binding subunit)"
/protein_id="AAW79214.1"
/db_xref="GI:21904336"
/translation="MTHFSKDPFVNMDLIFNQLMANMGYSRNPRLVNGREITPE
EFQIRQTGRQPLVATKATNSQMLTPKADSVLTQLNLTQEARQGHLDVPIGRNKEI
QDAEILARLNPVNLVGDAGVKGTAIVIEGLAQIYNGDVPAIAIKNEIVSIDSIL
EAGYQKGSFETQNLIOEVKEAGNIIIFDEIHOIVGAGATSSDSGSKGLAILKP
ALSREGTLIGATQDQENRNTILKNAALAREFNEVKVNPASAEPTFHLMGIRNLYEQ
HHHTLPDNLKAAVDYSIQVIPQRSPLDKAIDLLDMTAAHLAAQHPVTDLTLEI

```

```

AKOKESQKAVAKEDFEKALAAKTRITLQKIOEHQNSONTATVNDIAESVERLTG
IPVSNMGTNDLERLKGISSRLKSHVIGODERAAVAARAIERNRAGFDGRRPVSFLP
VGPTGKTELAKQLADLFGSKDAIIRLDMSEYNDRTAVSKLIGTAGTVGVGDDNNN
TLTERVRNRPYAIIVLDEIEKADPQIIITLLQLDDGLDQGGTINFRKNTVIATIS
NAGFGQDQTESNIMDRAPYFRPEFLNRFNSIIFKNHLOKESLEIEVDLLAEVN
QTTAKKGISLITDDAKAHLIDLCYNHANGARPLRIIEQIRDIRITDYVLDHPVKK
LOAILKEGOLVIRQNDQ"
7147. 7830
/gene="rpiA"
/note="Spym3_0608"
7147. 7830
/gene="rpiA"
/note="best blastp hit: gb|AAK33809.1| (AE006538) putative
ribose 5-phosphate isomerase [Streptococcus pyogenes M1
GAS]"
/codon_start=1
/transl_table=11
/product="putative ribose 5-phosphate isomerase"
/protein_id="AAW79215.1"
/db_xref="GI:21904337"
/translation="MEALKKIAGVTAQYVDTGMTIGLTGSGTAYYVEEIGRRVKOE
GLQVGVTTSSVTSKQAEVLGIPKSIDIDISLITVDGADEVKNGIKGGGAALL
MEKIVATPKETIYVVDASKWVHLAGFKLPVEVVOYGADRLFRVFKAGYKPSFRHK
GDSRLVDMQNYIIDLGLGCIKDPVAFGHLLDGTGVGVVEHGLFNGVMYKXIVASKDGV
TVLEAPTAG"
7907. 9118
/gene="deob"
/note="Spym3_0609"
7907. 9118
/gene="deob"
/note="best blastp hit: gb|AAK33810.1| (AE006538) putative
phosphopentomutase [Streptococcus pyogenes M1 GAS]"
/codon_start=1
/transl_table=11
/product="putative phosphopentomutase"
/protein_id="AAW79216.1"
/db_xref="GI:21904338"
/translation="MSKFNRIHLVLDVSGIAGAPADAKFEFNAGVADTSDTLCHISE
AAGLSVFNMAKIGLGNISRPILKTVPTEDNPTGIYTKLBEVSLGKDTMTGHWMEIGL
NITEPEDTFWNGPEEILKIEFSGRKKIIRANKPYSGTAVIDDFQPGDMETGELIV
YTSADPVQIAAHEDIIPEELYICEYARSITLERPALLGRIIARYPVGDGNGFTRT
ANRHDYAVSPFOOTVLNKADAGVPTVAVGKINDIFNGSGITNDMGHKNSHGIDTL
IKTLQLPEFTKGFSTNLVDFDANFGHRRDPESYRDLCHFEFDNRLPIIANMKEDLL

```

```

Query Match 19.6%; Score 114; DB 1; Length 53354;
Best Local Similarity 53.7%; Pred. No. 1.8e-18;
Matches 259; Conservative 0; Mismatches 220; Indels 3; Gaps 1;

Qy 1 ATGTTAATGCTTTTATTCATTATTTTGGCCTTACTGTAGGCAGCATTCCTTCGCTTA 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 22122 ATGAATTTACTACTTTTATTTACCATTGCGCTTACTTTTCTTCTTCTTCCAACTGGACTA 22063
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 61 ATTGTGGCAAGCTTGGCAAGGAATTCATATTCGGGACGACGGAAGCGCAACTTAGGC 120
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 22062 TGGATTGGACAGTACTTTTACCACATCACTTACGAGAGCATGGATCAGGAATACTGGA 22003
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 121 GCTACCAATGCTTCGTTACATTTGGTGTATAAGCTGTTTCGGTGTTCATAGCGGAGAT 180
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 22002 ACCACAAATACTTTTCGGATTTTAGGTGTCAAGCGCAGGAACAGCTTACCTAGCTATTGAT 21943
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 181 ATTTTGAAGGACACATGGCAACTGCATTTGCTTCTTCATGCATGT---TGATATTTCAC 237
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 21942 ATGTTTAAAGGACACTTTTCAATATTTGTTACCAATATTTTGGTATGACTTCAATTTCA 21883
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 238 CCGCTTCTTTCGAGGAGTCTTTTCGGTGTTCAGGCCACGTTTTCCTCATCTTCGCCAAATTT 297
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 21882 TCCATTGCTATCGGCTTTTTCGCAGTTTTCAGGCATACTTTTTCCTATTTTTCGCCAACTTT 21823
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 298 AAAGCGGTAAAGCGGTGGCGACATCAGGAGCGGTTTTCGATATTTTACGACACCCCTGTGA 357
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 21822 AAAGGTGTAAGGCGGTGACCAACAGTCTGCTGTATTGCTAGGCTTTGCTCCGTTATAT 21763
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 358 TTTATCAGATGGTTCGGGTATTCTTCATCTTTTATATCTTGCATAAATTTGTTTCTCTC 417
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Db 21762 CTCCTTTTATAGCATCTATCTTTGTTTATAGTCTCTATTTATTTAGCATGATATCTTTA 21703
Oy 418 TCATCGATGTTAAAGGATCTACTGTTATATATAGTCTTTGTTCCATGATACGTAT 477
Db 21702 GCTAGTGTGGTTCAGCTATCGTTGGTGTGTATCTGTTTAAACATTTCTCTGCCATTCAT 21643
Oy 478 TT 479
Db 21642 TT 21641

RESULT 14
AF269800/c
LOCUS AF269800 3014 bp DNA linear BCT 01-AUG-2000
DEFINITION Staphylococcus epidermidis strain SRI clone step.1023g03 genomic
sequence.
ACCESSION AF269800
VERSION AF269800.1 GI:9623698
KEYWORDS
SOURCE Staphylococcus epidermidis.
ORGANISM Staphylococcus epidermidis.
REFERENCE 1 (bases 1 to 3014)
Bacteria; Firmicutes; Bacillales; Staphylococcus.
AUTHORS Kimmerly,W.J., Taylor,J.David., Nelsen,A.J., Godlevski,M.M.,
Rubino,M.A., Nelson,F.J., Rivers,P.R., Torruella-Miller,I.,
Listenebee,S., Ashanti,C., Altschuller,G., Mamo,L., Shepherd,N.S.,
Fuchs,R., Fleming,T., Guan,X., Du,L., Cain,D.H., Miller,G.S. and
Furdon,P.J.
TITLE Transposon-mediated sequencing of the Staphylococcus epidermidis
genome
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3014)
Taylor,J.David., Kimmerly,W.J., Nelsen,A.J., Godlevski,M.M.,
Rubino,M.A., Nelson,F.J., Rivers,P.R., Torruella-Miller,I.,
Listenebee,S., Ashanti,C., Altschuller,G., Mamo,L., Shepherd,N.S.,
Fuchs,R., Fleming,T., Guan,X., Du,L., Cain,D.H., Miller,G.S. and
Furdon,P.J.
TITLE Direct Submission
JOURNAL Submitted (22-MAY-2000) Departments of Genomic Sciences and
Bioinformatics, Genetics Directorate, Glaxo Wellcome, Inc., 5 Moore
Drive, Research Triangle Park, North Carolina 27709-3398, USA
FEATURES
source
1. 3014
/organism="Staphylococcus epidermidis"
/strain="SRI"
/db_xref="taxon:1282"
/clone="step.1023g03"
BASE COUNT 1083 a 444 c 510 g 977 t
ORIGIN

Query Match 18.68; Score 108.2; DB 1; Length 3014;
Best Local Similarity 50.78; Pred. No. 5.7e-17;
Matches 309; Conservative 0; Mismatches 273; Indels 27; Gaps 1;

Oy 1 ATGTTAATTGCTTTTATGATTTATTTGGCTTCTGATAGGACGATCCATCTGGCTTA 60
Db 1223 ATGATGATCATCGTCATGTTAATCTTGAGTTATCTGATTTGCCAAGGGGTTA 1164
Oy 61 ATTGTGGCAAGCTTCCCAAGGAATTGATATTCGGGAGCACGGAAGCGCAACTTAGGC 120
Db 1163 ATTATTGGTAAATTTATTTTAAAGAGATATAAGACAATACGGTAGTGGAAATCTGGA 1104
Oy 121 GCTACCAATGCAATCCGTCATGTTGGTGTAAAGCTGTTGCGTTCATAGCCGGAGAT 180
Db 1103 GCAACTAACAGTTTTTCGTGTTCTTGGAAAGACGAGCTGGATTTATAGTTTACGTTT 1044
Oy 181 ATTTTGAAGGACACTGGCACTGCATTCCTTTCTCATGCTGTTGATTTACACCG 240
Db 1043 ATTTTCAGGGATTTATTACAGTCTTTTCCACTATGTTCCAGTTTCATCGCGATGGT 984
Oy 241 CTTCCTT-----GCAGGAGTCTTTGCGGTTTTTAGGCCAC 273
Db 983 GTTATAAGCACCTTCTTTTACAAATGGTTTAAATAGTAGGATTTGTGCAATACTCGTCCAC 924

Db 983 GTTATAAGCACCTTCTTTTACAAATGGTTTAAATAGTAGGATTTGTGCAATACTCGTCCAC 924
Oy 274 GTGTTTCCCATCTTCGCCCAATTTAAAGCGGTAAAGCCGTGGCGACATCAGGAGCGTT 333
Db 923 GTGTATCCAATATATATCGAAATTTAATGGCGGAAAGACAGTAGTACCAAGTGCAGGAGTT 864
Oy 334 TTGCTATTTTACGCACCCCTGTTATTATATACAGATGGTTGCGGTATTTCTTCATCTTTTAA 393
Db 863 GTATTAGGTGCAATCTCTATTTTACTTCTTATCTTGCAATATATCTTTTATAGTATTA 804
Oy 394 TACTTGACTAAATTTGTTCTCTCTCATCGATGTTACAGGGATCTATACTGTTATATAT 453
Db 803 AATATCTTTAAATATGTTCTTTTATCAAGTATCATTCGACCAATAGTGTGCTGATGGT 744
Oy 454 AGTTTCTTGTCCATGATACGTATTTATGATTCGTTACCTGCTACCTGCTACCTATTTTGTG 513
Db 743 TCAATCATCATTCATGATATATATTTTACTTCTGCTGTAGCGGAATTTGTTCAATCATATTA 684
Oy 514 ATATACAGACACCGGAGCAACATTAAAGAAATATCAATAAAGACAGACCTTAAAGTAA 573
Db 683 ATPATTCACACAATCTAATATAGTTAGATTTTAAAGGAGAGAACCTTAAATTA 624
Oy 574 TGGTTATAA 582
Db 623 TGGATGTAA 615

RESULT 15
AX145118/c
LOCUS AX145118 3014 bp DNA linear PAT 31-MAY-2001
DEFINITION Sequence 3840 from Patent WO0134809.
ACCESSION AX145118
VERSION AX145118.1 GI:14283683
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
REFERENCE 1 (bases 1 to 3014)
AUTHORS Kimmerly,W.J.
TITLE Staphylococcus epidermidis nucleic acids and proteins
JOURNAL Patent: WO 0134809-A 3840 17-MAY-2001;
GLAXO GROUP LIMITED (GB)
FEATURES
source
1. 3014
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="synthetic nucleic acid sequence"
BASE COUNT 1083 a 444 c 510 g 977 t
ORIGIN

Query Match 18.68; Score 108.2; DB 6; Length 3014;
Best Local Similarity 50.78; Pred. No. 5.7e-17;
Matches 309; Conservative 0; Mismatches 273; Indels 27; Gaps 1;

Oy 1 ATGTTAATTGCTTTTATGATTTATTTGGCTTCTGATAGGACGATCCATCTGGCTTA 60
Db 1223 ATGATGATCATCGTCATGTTAATCTTGAGTTATCTGATTTGCCAAGGGGTTA 1164
Oy 61 ATTGTGGCAAGCTTGGCAAGGAATTGATATTCGGGAGCACGGAAGCGCAACTTAGGC 120
Db 1163 ATTATTGGTAAATTTATTTTAAAGAGATATAAGACAATACGGTAGTGGAAATCTGGA 1104
Oy 121 GCTACCAATGCAATCCGTCATGTTGGTGTAAAGCTGTTGCGTTCATAGCCGGAGAT 180
Db 1103 GCAACTAACAGTTTTTCGTGTTCTTGGAAAGACGAGCTGGATTTATAGTTTACGTTT 1044
Oy 181 ATTTTGAAGGACACTGGCACTGCATTCCTTTCTCATGCTGTTGATTTACACCG 240
Db 1043 ATTTTCAGGGATTTATTACAGTCTTTTCCACTATGTTCCAGTTTCATCGCGATGGT 984
Oy 241 CTTCCTT-----GCAGGAGTCTTTGCGGTTTTTAGGCCAC 273
Db 983 GTTATAAGCACCTTCTTTTACAAATGGTTTAAATAGTAGGATTTGTGCAATACTCGTCCAC 924

[illegible]

Search completed: December 25, 2002, 10:52:49
Job time : 10757.6 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 24, 2002, 19:31:45 ; Search time 136.703 seconds
(without alignments)
9587.638 Million cell updates/sec

Title: US-10-068-080-4

Perfect score: 582
Sequence: 1 atgttaattgctttattgat.....ctaaagtaaaatggttataa 582

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:*
- 24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	582	100.0	582	20	AAZ20371
2	582	100.0	582	21	AAZ09182
3	176	30.2	495269	24	ABQ67195
4	117.4	20.2	636	24	ABN67449
5	115.8	19.9	621	24	ABN70980
6	114	19.6	720	24	ABN67450
c 7	109.6	18.8	6691	20	AAZ13025
8	108.2	18.6	615	24	ABN92939
c 9	108.2	18.6	3014	22	AAH54476

c	10	108.2	18.6	4105	22	AAH54115	S. epidermidis gen.
	11	106.6	18.3	594	22	AAH53731	S. epidermidis ope
	12	101.6	17.5	642	23	AAH55594	Streptococcus pneu
	13	101.6	17.5	642	23	AAH55801	Streptococcus pneu
	14	100	17.2	642	21	AAA09181	S. pneumoniae S-yn
	15	100	17.2	6171	19	AAV52170	Streptococcus pneu
	16	90.6	15.6	609	23	AAH54402	Streptococcus pneu
	17	88.2	15.2	606	23	AAH51668	Staphylococcus aur
c	18	80.4	13.8	818	18	AAV75059	Staphylococcus aur
	19	76.6	13.2	344	20	AAZ20352	Gene encoding bact
c	20	55.2	9.5	1657	24	ABQ70507	Listeria monocytog
	21	48	8.2	600	23	AAH53263	Haemophilus influe
	22	48	8.2	1830121	17	AAZ42063	Haemophilus influe
	23	46.4	8.0	72	24	ABK76809	Bacillus lichenifo
c	24	45.2	7.8	556	23	ABV40063	Human prostate exp
c	25	45.2	7.8	556	23	ABV40163	Human prostate exp
	26	45.2	7.8	556	23	ABV42105	Human prostate exp
c	27	45.2	7.8	556	23	ABV43601	Human prostate exp
	28	43.4	7.5	648	24	ABQ48108	Oligonucleotide fo
c	29	43.4	7.5	648	24	ABQ48109	Oligonucleotide fo
	30	43	7.4	789	23	AAH53719	Helicobacter pylor
	31	43	7.4	823	19	AAZ14476	H. pylori GHPO 108
c	32	42.6	7.3	309	23	ABV44994	Human prostate exp
	33	42.4	7.3	612	23	AAH56022	Human prostate exp
	34	42.2	7.3	10205	24	ABL70236	Salmonella typhi D
	35	42.2	7.3	10205	24	ABL31275	Chemically treated
	36	42	7.2	9905	24	ABL32062	Signal transductio
	37	41.8	7.2	6956	24	ABL70225	Human immune syste
	38	40.8	7.0	596	22	AAZ14782	Chemically treated
	39	40.2	6.9	626	24	ABQ56717	Human breast cancer
	40	40.2	6.9	752	24	ABQ14854	Human colon cancer
c	41	40.2	6.9	752	24	ABQ14855	Oligonucleotide fo
	42	39.8	6.8	663	23	AAH53880	Oligonucleotide fo
c	43	39.6	6.8	394	22	AAZ10020	Helicobacter pylor
c	44	39.6	6.8	476	23	ABV58368	Human breast cancer
	45	39.6	6.8	6121	24	ABL92201	Human prostate exp
							Chemically treated

ALIGNMENTS

RESULT 1

AAZ20371
ID AAZ20371 standard; DNA; 582 BP.

AC AAZ20371;

XX 17-NOV-1999 (first entry)

DE B. subtilis B-ynes protein coding sequence.

XX General essential protein; pathogenic bacteria; pathogen; inhibitor;

KW bacterial growth; B-ynes; ds.

XX Bacillus subtilis.

XX Key Location/Qualifiers

FT CDS 1..582

FT /*tag= a

FT /product= B-ynes

XX WO9933871-A2.

XX 08-JUL-1999.

XX 30-DEC-1998; 98WO-0527918.

XX 31-DEC-1997; 97US-0070116.

XX (MILL-) MILLENNIUM PHARM INC.

XX Youngman P, Fritz C, Murphy C, Guzman L;

XX

S. epidermidis gen-
S. epidermidis ope
Streptococcus pneu
Streptococcus pneu
S. pneumoniae S-yn
Streptococcus pneu
Staphylococcus aur
Staphylococcus aur
Gene encoding bact
Listeria monocytog
Haemophilus influe
Bacillus lichenifo
Human prostate exp
Human prostate exp
Human prostate exp
Human prostate exp
Human prostate exp
Oligonucleotide fo
Oligonucleotide fo
Helicobacter pylor
H. pylori GHPQ 108
Human prostate exp
Salmonella typhi D
Chemically treated
Signal transductio
Human immune syste
Chemically treated
Human breast cance
Human colon cancer
Oligonucleotide fo
Oligonucleotide fo
Helicobacter pylor
Human breast cance
Human prostate exp
Chemically treated

DR WPI: 1999-430230/36.
DR P-PSDB: AAY22580.
XX Streptococcus pneumoniae general essential protein genes and proteins,
PT useful for identification of antibacterial agents
XX
XX Disclosure; Fig 24; 124pp; English.
XX
CC This sequence encodes the Bacillus subtilis B-ynes protein. B-ynes is
CC related to the Streptococcus pneumoniae general essential
CC protein (GEP) gene of the invention. The genes encoding the GEP
CC polypeptides are useful molecular tools for identifying similar genes in
CC pathogenic microorganisms, such as pathogenic strains of Bacillus. In
CC addition, the operons containing genes encoding GEP and the polypeptides
CC themselves, are useful targets for identifying compounds that are
CC inhibitors of the pathogens in which the GEP are expressed. Such
CC inhibitors are useful for inhibiting bacterial growth by being
CC bacteriostatic or bacteriocidal.
XX
XX Sequence 582 BP; 147 A; 115 C; 122 G; 198 T; 0 other;
SQ
Query Match 100.0%; Score 582; DB 20; Length 582;
Best Local Similarity 100.0%; Pred. No. 1.6e-158;
Matches 582; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ATGTTAATTCGCTTATTGATTTATTTGGCCTACTTGTATAGGCACGATTCCTCTGGCTTA 60
Db 1 ATGTTAATTCGCTTATTGATTTATTTGGCCTACTTGTATAGGCACGATTCCTCTGGCTTA 60
Qy 61 ATTGTGGCAAGCTTGGCAAGGAATTGATATTCGGGAGCACGGAAGCGCACTTAGGC 120
Db 61 ATTGTGGCAAGCTTGGCAAGGAATTGATATTCGGGAGCACGGAAGCGCACTTAGGC 120
Qy 121 GCTACCAATGCTTCCGTACATTCGGGTGTAAGAGCTGGTTCGTCATAGCCGGAGAT 180
Db 121 GCTACCAATGCTTCCGTACATTCGGGTGTAAGAGCTGGTTCGTCATAGCCGGAGAT 180
Qy 181 ATTTTGAAGGACACATGGCAACTGCATTCGCTTTCTCATGATGTTGATATTCACCCG 240
Db 181 ATTTTGAAGGACACATGGCAACTGCATTCGCTTTCTCATGATGTTGATATTCACCCG 240
Qy 241 CTCTTCGAGGAGCTTTTGGCGTTTATAGCCACGCTGTTTCCCATCTTCGCCAAATTTAAA 300
Db 241 CTCTTCGAGGAGCTTTTGGCGTTTATAGCCACGCTGTTTCCCATCTTCGCCAAATTTAAA 300
Qy 301 GCGGTAAGCCGTTGGCGACATCAGAGGCGTTTGGCTATTTTACGACCCCTCTTATT 360
Db 301 GCGGTAAGCCGTTGGCGACATCAGAGGCGTTTGGCTATTTTACGACCCCTCTTATT 360
Qy 361 ATCAGATGTTGCGGTATTCCTCATCTTTTATATATAGTTTCTCTCTCTCA 420
Db 361 ATCAGATGTTGCGGTATTCCTCATCTTTTATATATAGTTTCTCTCTCTCA 420
Qy 421 TCGATGTTTAAACAGGATCTATACCTGTTATATATAGTTTCTTGTCCATGATACATTTA 480
Db 421 TCGATGTTTAAACAGGATCTATACCTGTTATATATAGTTTCTTGTCCATGATACATTTA 480
Qy 481 TTGATTCGTTACCTGCTCCTATTTTGTGATATACAGACACCGAGCAATTTAAA 540
Db 481 TTGATTCGTTACCTGCTCCTATTTTGTGATATACAGACACCGAGCAATTTAAA 540
Qy 541 CGAATTATCAATAAACAAGACCTTAAAGTAAATGGTTATATA 582
Db 541 CGAATTATCAATAAACAAGACCTTAAAGTAAATGGTTATATA 582
RESULT 2
AAA09182
ID AAA09182 standard; DNA: 582 BP.
XX
AC AAA09182;
XX
DT 10-AUG-2000 (first entry)

XX B. subtilis B-ynes coding sequence.
DE
XX
XX B-ynes; S-ynes; survival; antibacterial; inhibitor; ds..
KW
XX Bacillus subtilis.
OS
XX
XX WO200020527-A1.
PN
XX
XX 13-APR-2000.
PD
XX
XX 30-SEP-1999; 99WO-US22665.
PF
XX
XX 30-SEP-1998; 98US-0163445.
PR
XX
XX (MILL-) MILLENNIUM PHARM INC.
PA
XX
XX Fritz C, Youngman P, Guzman L;
PI
XX
XX WPI: 2000-303799/26.
DR
XX
XX P-PSDB: AAY92247.
Methods for identifying an antibacterial agent for treating
Streptococcus pneumoniae infections comprises detecting an interaction
PT between a ynes polypeptide and a test compound
PT
XX
XX Disclosure; Fig 2; 65pp; English.
PS
XX
XX Bacillus subtilis B-ynes is a homologue of the Streptococcus pneumoniae
CC ynes gene (S-ynes) which is essential for survival for a wide range of
CC bacteria. Identifying an antibacterial agent comprises contacting a ynes
CC polypeptide (S-ynes) with a test compound and detecting an interaction of
CC the test compound with the S-ynes polypeptide which indicates that the
CC compound is an antibacterial agent. Alternatively, detecting a decrease
CC in function of the polypeptide contacted with the test compound and
CC determining whether the compound inhibits growth of bacteria, relative to
CC the growth of bacteria cultured in the absence of a test compound where
CC inhibition of growth indicates the compound is an antibacterial agent.
CC Inhibitors of S-ynes function are useful for treating a S. pneumoniae
CC infection in mammals.
XX
XX Sequence 582 BP; 147 A; 115 C; 122 G; 198 T; 0 other;
SQ
Query Match 100.0%; Score 582; DB 21; Length 582;
Best Local Similarity 100.0%; Pred. No. 1.6e-158;
Matches 582; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 ATGTTAATTCGCTTATTGATTTATTTGGCCTACTTGTATAGGCACGATTCCTCTGGCTTA 60
Db 1 ATGTTAATTCGCTTATTGATTTATTTGGCCTACTTGTATAGGCACGATTCCTCTGGCTTA 60
Qy 61 ATTGTGGCAAGCTTGGCAAGGAATTGATATTCGGGAGCACGGAAGCGCACTTAGGC 120
Db 61 ATTGTGGCAAGCTTGGCAAGGAATTGATATTCGGGAGCACGGAAGCGCACTTAGGC 120
Qy 121 GCTACCAATGCTTCCGTACATTCGGGTGTAAGAGCTGGTTCGTCATAGCCGGAGAT 180
Db 121 GCTACCAATGCTTCCGTACATTCGGGTGTAAGAGCTGGTTCGTCATAGCCGGAGAT 180
Qy 181 ATTTTGAAGGACACATGGCAACTGCATTCGCTTTCTCATGATGTTGATATTCACCCG 240
Db 181 ATTTTGAAGGACACATGGCAACTGCATTCGCTTTCTCATGATGTTGATATTCACCCG 240
Qy 241 CTCTTCGAGGAGCTTTTGGCGTTTATAGCCACGCTGTTTCCCATCTTCGCCAAATTTAAA 300
Db 241 CTCTTCGAGGAGCTTTTGGCGTTTATAGCCACGCTGTTTCCCATCTTCGCCAAATTTAAA 300
Qy 301 GCGGTAAGCCGTTGGCGACATCAGAGGCGTTTGGCTATTTTACGACCCCTCTTATT 360
Db 301 GCGGTAAGCCGTTGGCGACATCAGAGGCGTTTGGCTATTTTACGACCCCTCTTATT 360
Qy 361 ATCAGATGTTGCGGTATTCCTCATCTTTTATATATAGTTTCTCTCTCTCA 420
Db 361 ATCAGATGTTGCGGTATTCCTCATCTTTTATATATAGTTTCTCTCTCTCA 420
Qy 421 TCGATGTTTAAACAGGATCTATACCTGTTATATATAGTTTCTTGTCCATGATACATTTA 480
Db 421 TCGATGTTTAAACAGGATCTATACCTGTTATATATAGTTTCTTGTCCATGATACATTTA 480
Qy 481 TTGATTCGTTACCTGCTCCTATTTTGTGATATACAGACACCGAGCAATTTAAA 540
Db 481 TTGATTCGTTACCTGCTCCTATTTTGTGATATACAGACACCGAGCAATTTAAA 540
Qy 541 CGAATTATCAATAAACAAGACCTTAAAGTAAATGGTTATATA 582
Db 541 CGAATTATCAATAAACAAGACCTTAAAGTAAATGGTTATATA 582

ID	ABN70980 standard; DNA; 621 BP.
XX	
XX	ABN70980;
AC	
XX	
XX	01-JUL-2002 (first entry)
DT	
DT	
XX	
XX	Streptococcus polynucleotide SEQ ID NO 9873.
DE	
XX	
XX	Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW	group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
KW	antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
XX	
XX	Streptococcus agalactiae.
OS	
XX	
XX	WO200234771-A2.
PN	
XX	
XX	02-MAY-2002.
PD	
XX	
XX	29-OCT-2001; 2001WO-GB04789.
PF	
XX	
XX	27-OCT-2000; 2000GB-0026333.
PR	
PR	24-NOV-2000; 2000GB-0028727.
PR	07-MAR-2001; 2001GB-0005640.
PR	

XX
PA (CHIR-) CHIRON SPA.
PA (GENO-) INST GENOMIC RES.
XX
PI Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;
PI Tettelin H;

[illegible]

137 GTACATTGGGTGTAAGCTGGTTCGGTCGTCATAGCCGGAGATATTTGAAGGGACAC 196

137 GTACATTGGGTGTAAGCTGGTTCGGTCGTCATAGCGGAGATATTTGAAGGGACAC 196

Db 122 GTATTTAGGTGTAAGAGCTGGCAATGTCATTTGACTATTTGACATTTTAAAGGGGACTC 181
Qy 197 TGGCAACTGCAATGCTCTTCTCATGATGTTGATA---TTCACCCGCTCTTTCGAGGAG 253
Db 182 TAGCAACACTATTCCTATCATATAGGTATAGGATACGACAGATCTCCATTTTATCGGTT 241
Qy 254 TCTTTGGGTTTATAGGCCAGCGTGTTCCTCCATCTTCGCCAAATTTAAAGGCGGTAAGCCG 313
Db 242 TTTTGTGCTATATAGGTACACACTTTCCTATTTTTCGACAAATTCAAAGGTGGGAAGCCG 301
Qy 314 TGGCGCATCAGGAGGCGTTTGTCTATTTTACGACACCCCTGTTATTTATACGATGTTG 373
Db 302 TTGCTACAAGCGCTGGGGTTTGTAGGATTTGCACCTTCTTTTCTTCTTACCTTATAG 361
Qy 374 CGGTATCTTCATCTTTTATACCTGACTAAATTTGTTCTCTCATCGATGTTACAG 433
Db 362 TTATCTTTTATTAACGCTTTATCTTTTATGACATGATTTCCCTATCCAGTATTTACAGTAG 421
Qy 434 -GGATCTATACCTGTTATATATAGTTTCTTTGTCATGATAGCTATTTAT-----TGATTG 487
Db 422 CTGTTGTAGGTATTTCTTAGTGTCTTAATTTTCCATTAGTAGGCTTTATATATTAACGGACT 481
Qy 488 TCGTTACCTGCTCAGTATTTTGTGATATACAGACACCGAGCGCAACATTTAAAGCAATTA 547
Db 482 ACGACTGGATATTTTACCACTGTGTTATCTTAATGCGCCTAACGATTTATTCGAGATC 541
Qy 548 TCAATAAAGAGAACCTAAAGTAAATGTTTATAA 582
Db 542 AGGATAATATCAAAAGCTTATCGGAAGAGGCAAGAA 576

RESULT 6

ABN67450
ID ABN67450 standard; DNA; 720 BP.
XX
AC ABN67450;
XX
DT 01-JUL-2002 (first entry)
XX
DE Streptococcus polynucleotide SEQ ID NO 2813.
XX
KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
KW antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
OS Streptococcus pyogenes.

PN WO200234771-A2.

XX
XX
PD 02-MAY-2002.XX
XX
PF 29-OCT-2001; 2001WO-GB04789.XX
XX
PR 27-OCT-2000; 2000GB-0026333.XX
XX
PR 24-NOV-2000; 2000GB-0028727.XX
XX
PR 07-MAR-2001; 2001GB-0005640.XX
XX
PA (CHIR-) CHIRON SPA.XX
XX
PA (GENO-) INST GENOMIC RES.XX
XX
PI Telford J, Massignani V, Margarit Ros YI, Grandi G, Fraser C;XX
XX
PI Tettelin H;XX
XX
XX WPI: 2002-352536/38.XX
XX
DR P-PSDB: ABP26819.XX
XX
XXXX
XX
XXXX
XX
XXXX
XX
XXXX
XX
XXXX
XX
XXXX
XX
XXXX
XX
XXXX
XX
XXXX
XX
XXXX
XX
XXXX
XX
XXXX
XX
XXXX
XX
XXXX
XX
XXXX
XX
XX

CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in
CC the specification. The proteins have antibacterial and antiinflammatory
CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC antibodies that bind (I) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC Nucleic acids encoding (I) are used to detect Streptococcus in a
CC biological sample. (I) is used to determine whether a compound binds to
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC used as a vaccine or diagnostic composition. The disease caused by
CC Streptococcus that is prevented or treated may be meningitis. Nucleic
CC acid encoding (I) may be used to recombinantly produce (I) and may be
CC used in gene therapy. Antibodies to (I) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
CC Streptococcus proteins.

XX
SQ Sequence 720 BP; 199 A; 123 C; 112 G; 286 T; 0 other;

Query Match 19.6%; Score 114; DB 24; Length 720;

Best Local Similarity 53.7%; Pred. No. 6e-23;

Matches 259; Conservative 0; Mismatches 220; Indels 3; Gaps 1;

Qy 1 ATCTTAATGCTTTTATTTGATTTATTTGGCTACTTTGATAGGCAGCATTTCCATCTGGCTTA 60

Db 82 ATGAATTTACTACTTTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 141

Qy 61 ATTGTGGGCAAGCTTGGCAAGGAATTTGATTTTTCGGAGACGGAACGGCACTTAGG 120

Db 142 TGGATTGGCAGAGTACTTTTACCAATCACTTACGAGAGCATGGATCAGGAATTTACTGGA 201

Qy 121 GCTACCAATGCTTCCGTACATTTGGGTGTAAAGCTGGTTCGGTCTGTCATAGCCGGAGAT 180

Db 202 ACCACAAATTTCTTTCGGATTTTAGGTGTCAAGCAGGAACAGCTTACCTTAGCTATTGAT 261

Qy 181 ATTTGAAAGGACACTGGCAACTGCATTTGCTTTTCTCATGCTGT---TGATATTTCAC 237

Db 262 ATGTTTAAAGGACACTTTTCAATATTCTTACCAATTTATTTTGTATGACTTCAATTTCA 321

Qy 238 CCGTCTCTTCGAGAGTCTTTTGGGTTTATAGGCACGCTTTTCCCATCTTCGCCAAATTT 297

Db 322 TCCATTCTCTATCGGCTTTTTCGGAGTTTATAGGCATACTTTTCTTATTTTGGCAACTTT 381

Qy 298 AAAGCGGTAAAGCGTGGCGACATCAGAGGCGCTTTTGTCTATTTTACGACCCCTGTTA 357

Db 382 AAAGTGTGTAGGCGGTAGCAACAAAGTCTGGTGTATTTGCTAGGCTTTGCTCCGTTATAT 441

Qy 358 TTTATCAGATGTTGCGGTATTTCTTCATCTTTTATTTTATTTTATTTTATTTTCTCTC 417

Db 442 CTCCTTTTATAGCATCTATCTTTTGTGTTTATTTTATTTTATTTTATTTTATTTTATTT 501

Qy 418 TCATCGATGTTTACAGGATCTATATCTGTTATATATATAGTTTCTTTGTCATGATACGAT 477

Db 502 GCTAGTGTGTTTTCAGCTATCGTTGGTGTGTATCTGTTTATTAACATTTTCTCGCATCAT 561

Qy 478 TT 479

Db 562 TT 563

RESULT 7

AAI13025/C

ID AAI13025 standard; DNA; 6691 BP.

XX
XX
AC AAI13025;XX
XX
XX

DT 19-MAR-1999 (first entry)

XX
XX
XX Enterococcus faecalis genome contig SEQ ID NO:88.XX
XX
XX Enterococcus faecalis; contig; detection; Enterococcal infection;XX
XX
XX vaccine; attenuation; computer readable medium; ds.XX
XX
XXXX
XX
XXXX
XX
XXXX
XX
XXXX
XX
XXXX
XX
XXXX
XX
XXXX
XX
XXXX
XX
XX

OS Enterococcus faecalis.

PN WO9850555-A2.

PD 12-NOV-1998.

PF 04-MAY-1998; 98WO-US08985.

PX 14-NOV-1997; 97US-0066009.

PR 06-MAY-1997; 97US-0044031.

PX 16-MAY-1997; 97US-0046655.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Barash SC, Dillon PJ, Kunsch CA;

PX WPI; 1999-045171/04.

PT New isolated Enterococcus faecalis polynucleotides and polypeptides
PT - used to develop products for the detection of Enterococcus and for
PT use in vaccines for prevention or attenuation of Enterococcus
PT infection.

PS Claim 1; Page 592-595; 2084pp; English.

CC A computer readable medium has been developed which has recorded on it
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC AAX12938 to AAX13919 represent these nucleotide sequences which are
CC primary nucleotide sequences, also known as contigs. The computer-based
CC system can identify fragments of the Enterococcus faecalis genome with
CC commercial importance. The products can be used to detect the presence
CC of Enterococcus faecalis in samples. They can also be used for
CC diagnosing Enterococcal infection in an animal and monitoring
CC progression of disease, and for identifying agents which can be used to
CC modulate the growth or pathogenicity of Enterococcus faecalis, or
CC another related organism, in vivo or in vitro. In particular the
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC can be used in vaccines to prevent or attenuate an Enterococcal
CC infection.

XX Sequence 6691 BP; 2383 A; 1085 C; 1355 G; 1858 T; 10 other;

Query Match 18.8%; Score 109.6; DB 20; Length 6691;
Best Local Similarity 53.4%; Pred. No. 2.7e-21;
Matches 254; Conservative 0; Mismatches 219; Indels 3; Gaps 1;

QY 1 ATGTTAATTCCTTATTTGATTTATTTGGCTTACTTGTATAGGACGATTCATCTGGCTTA 60

DB 5424 ATGAAATCGTCATTTTGTACTTGTGCTTATTTATTTAGTTCGATTCCTCCAGGTGTT 5365

QY 61 ATTGTGGGCAAGCTTCCCAAGGANTTGATATTCGGGAGCACGGAAGCGCACTTAGGC 120

DB 5364 TGGATTGGTAAACCTTTCTTTTAAAGAGATATACGCCAATTTGGGAGTGGGAATACAGGA 5305

QY 121 GCTACCAATCGATTCGTCATACATTTGGGTGTAAGCTGGTTCGTCATAGCGGAGAT 180

DB 5304 AACACCAATACATTCGTGCTTAGGGAACCTGCCGGAATTTACGGTATTTAATTCGAT 5245

QY 181 ATTTTGAAGGACACTGGCAACTGCATTCGCTTTTCTCATGCATGT---TGATATTAC 237.

DB 5244 ATCTTGAAGAACGTTAGCCACTTCATTACCCCTATTGTTGTTTACAAAGCGGTGAAT 5185

QY 238 CCGCTTCTTCAGAGTCTTTGGGGTTTATAGGCGACGTGTTCCCATCTTCGCGCAATTT 297

DB 5184 CCGCTCTTCTTTGGGTAGCAGCTGTTTATAGGCATACCTTCCTTATTTTGGCAATTC 5125

QY 298 AAGGGCGGTAAAGCCGTGGCGACATCAGGAGCGTTTGTGATATTTACGACCCCTGTTA 357

DB 5124 AAGGTGGTAAAGCCGTAGCCACTAGCGTGGCATGTTATATGATACAGACCCCATTT 5065

QY 358 TTTATCAGATGTTGCGGTATTTCTTCATCTTTTATATCTTACTAAATTTGTTTCTCTC 417

DB 5064 TTTATTTATCTGCTCTTATTTTGTGATTTGTCTGATCTACCAAGTATGGTGTGATTTA 5005

QY 418 TCATCGATGTTAACAGGGATCTACTGTATATATAGTTCTTTGTCATGATAC 473
DB 5004 ACAAGTATGATTAGCGCTGTACTAATTACACTTTCTTACTATTATTTGGCTTTAC 4949

RESULT 8

ABN92939

ID ABN92939 standard; DNA; 615 BP.

XX AC ABN92939;

XX DT 24-JUL-2002 (first entry)

XX Staphylococcus epidermidis ORF nucleic acid sequence SEQ ID NO:2402.

KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
KW antibacterial; gene therapy; gene; ds.

XX Staphylococcus epidermidis.

XX US6380370-B1.

XX PD 30-APR-2002.

XX PF 13-AUG-1998; 98US-0134001.

XX PR 14-AUG-1997; 97US-055779P.

XX PR 08-NOV-1997; 97US-064964P.

XX (GENO-) GENOME THERAPEUTICS CORP.

XX PI Doucette-Stamm LA, Bush D;

XX WPI; 2002-381255/41.

XX DR P-PSDB; ABP40394.

XX Novel isolated nucleic acid encoding a Staphylococcus epidermidis

PT polypeptide, useful for diagnosing and treating bacterial infections -

XX Disclosure; SEQ ID 2402; 267pp; English.

CC ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
CC frame (OFF) nucleic acid sequences which encode the amino acid sequences
CC given in ABP3124 to ABP3960. The S. epidermidis sequences have
CC antibacterial activity and can be used in gene therapy. The sequences
CC can also be used in the diagnosis and treatment of bacterial infections,
CC particularly S. epidermidis infections. The sequences can be used to
CC screen for compounds able to interfere with the S. epidermidis life
CC cycle or inhibit S. epidermidis infection.

CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC USPTO web site.

XX Sequence 615 BP; 185 A; 81 C; 112 G; 237 T; 0 other;

Query Match 18.6%; Score 108.2; DB 24; Length 615;
Best Local Similarity 50.7%; Pred. No. 2.7e-21;
Matches 309; Conservative 0; Mismatches 273; Indels 27; Gaps 1;

QY 1 ATGTTAATTCCTTATTTGATTTATTTGGCTTACTTGTATAGGACGATTCATCTGGCTTA 60

DB 7 ATGATGATCATCGTCATGTTAATCTTTGAGTTATCTGATTTCCCAAGCGGTTA 66

QY 61 ATTGTGGCAAGCTTCCCAAGGAATTTGATATTCGGGAGCACGGAAGCGCAACTTAGGC 120

DB 67 ATTAATGTTGAATATTTTAAAAAGATATAAGACATACGTTAGTGGAAATCTGGA 126

QY 121 GTTACCAATGCAATTCGTTACATTTGGGTGTAAAGCTGGTTCGTCATACCCGAGAT 180

DB 127 GCAACTAACAGTCTTCGTTCTTGGAGACAGCTGGATTTATAGTTACGTTTTTATAGAT 186

QY 181 ATTTTGAAGGACACTGCAACTGCAATTCCTTCTCATGCAATGTTGATATTCACCCG 240

[illegible]

RESULT 9	
AAH54476/c	
ID	AAH54476 standard; DNA; 3014 BP.
XX	
XX	AAH54476;
XX	
XX	03-SEP-2001 (first entry)
DT	
XX	
DE	S. epidermidis genomic polynucleotide sequence SEQ ID NO:3840.
XX	
KW	Staphylococcus epidermidis SRI strain; infection; diagnosis;
KW	vaccination; endocarditis; ds.
XX	
OS	Staphylococcus epidermidis.
XX	
PN	WO200134809-A2.
XX	
PD	17-MAY-2001.
XX	
PP	09-NOV-2000; 2000WO-US30782.
XX	
PR	09-NOV-1999; 99US-0164258.
XX	
PA	(GLAX) GLAXO GROUP LTD.
XX	
PI	Kimmerly WJ;
XX	
DR	WPI; 2001-316495/33.
XX	
PT	Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
PT	useful for vaccinating against infections, e.g. endocarditis -
XX	
PS	Claim 8; Page 1482-1483; 2188pp; English.
XX	
CC	AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC	(II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
CC	(I) and (II) can have antibacterial activity and therefore can be used
CC	in vaccination. The nucleic acids (I) may be used to produce the
CC	S. epidermidis polypeptides (II) via the production of vectors
CC	containing them which are used to produce hosts cells which express the
CC	polypeptides. The polypeptides (II) (and/or nucleic acids) may then be

CC	used to vaccinate subjects and to raise antibodies against the bacteria.
CC	The polypeptides may also be used to assay for other inhibitors of their
CC	activity and therefore identify compounds that may be used for the
CC	treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC	AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC	polynucleotide sequences from the present invention. AAH55091 to
CC	AAH55098 represent oligonucleotide sequences and primers which are used
CC	in the exemplification of the present invention.
CC	N.B. The present invention specifically claims all the polynucleotide
CC	sequences given in the sequence listing of the present specification,
CC	however the sequence listing only goes up to SEQ ID NO:4454 so even
CC	though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
CC	no sequences are present for SEQ ID NO:4455 to 4464.
XX	
SX	Sequence: 3014 BP; 1083 A; 444 C; 510 G; 977 T; 0 other;
	Query Match 18.6%; Score 108.2; DB 22; Length 3014;
	Best Local Similarity 50.7%; Pred. No. 5 le-21;
	Matches 309; Conservative 0; Mismatches 273; Indels 27; Gaps 1;
Qy	1 ATGCTTAATTCCTTTATTGTGATTAATTTTGGCCTACTTGATAGGCAGCATTTCCATCTGGCGTTA 60
Db	
	1223 ATGATGATCATCGTCATGTTAACTCTTCAGTATATCTGATTTGGTGCATTTCCCAAGCGGTTA 1164
Qy	61 ATTGTGGCGGAAGCTTGCCAAGGAATTTGATATTCGGGAGACACGGACGCAACTTAGGC 120
Db	
	1163 ATTATTCGGTAAATTAATTTTTAAAAAAGATAATAGAACAATACGGTAGTGGAAATACATGGA 1104
Qy	121 GCTACCAGATGCATCCGTFACATTCGGGTGTAAGAGCTGGTTCGGTGCATAGCCGGAGAT 180
Db	
	1103 GCACATAACAGTTTTCGTGTTCTTGGGAAGACCAGCTCGATTTATAGTTTACGTTTTAGAT 1044
Qy	181 ATTTTGAAGAGGACATGGCAACTGCATTCGCTTTTCTCATGCAATGTGATPATTCACCCG 240
Db	
	1043 ATTTTCAAGGAGTATTATTACAGTCTTTTCCACTATGTTGCCAGTTCATCGCGGATGGT 984
Qy	241 CTTCCTT-----GCAGGAGTCTTTGGCGTTTAGGCCAC 273
Db	
	983 GTTATAAGCACCTCTTTTACAAATGGTTTAATAGTAGGATGTTTTCGAATACTCGGTCAC 924
Qy	274 GTGTTTCCCCTCTTCGCCCAAATTTAAAGCGGTAAAGCGCGGCACATCAGGAGGGTT 333
Db	
	923 GTGATCAATATATCTGAAATTTAATGGCGGAAGACGATAGCTACCAAGTCAGGAGTT 864
Qy	334 TTGCTATTTTACGCACCCCTGTTATTTATCATCAGATGGTTCGGGTATTTCTTCATCTTTTA 393
Db	
	863 GTAATTAGGTGTCAAATCTATTTTACTTCTTATCTTGGCAATTAATCTTTTGTAGTGATTA 804
Qy	394 TACVTGACTAAATTTGTTTCTCTCATCGATGTTTACAGGGATCTATACGTGTTATATAT 453
Db	
	803 AAAATCTTTAAATATGTTTCTTTATCAAGTATCATTCGACGAATTAGTGTGTGATTTGGT 744
Qy	454 AGTCTCTTTGCGCATGATACGTATTTATGATTGCTGCTTACCCCTGCTCACTATTTTGTG 513
Db	
	743 TCATATCATCATCATGATATATTTTACTTGCTGTTAGCGGAATGTTTCAATCATATTA 684
Qy	514 ATATACAGACACCGAGGCAACATTTAAACGAATTAATCAATAAACACCTTAAAGCTAAA 573
Db	
	683 ATAATTCGACACAAATCTAATATAGTAGTAATTTTTAAAGGAGAAGAACCTTAAATATA 624
Qy	574 TGGTTATAA 582
Db	
	623 TGGATGTAA 615
RESULT 10	
AAH54115/c	
ID	AAH54115 standard; DNA; 4105 BP.
XX	
AC	AAH54115;
XX	
DT	03-SEP-2001 (first entry)
XX	

DE S. epidermidis genomic polynucleotide sequence SEQ ID NO:3479.
 XX Staphylococcus epidermidis SRI strain; infection; diagnosis;
 KW vaccination; endocarditis; ds.
 XX
 OS Staphylococcus epidermidis.
 PN WO200134809-A2.
 XX
 PD 17-MAY-2001.
 XX
 PF 09-NOV-2000; 2000WO-US30782.
 XX
 PR 09-NOV-1999; 99US-0164258.
 XX
 PA (GLAX) GLAXO GROUP LTD.
 XX
 PI Kimmerly WJ;
 XX
 DR WPI; 2001-316495/33.
 XX
 PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
 PT useful for vaccinating against infections, e.g. endocarditis .
 PS Claim 8; Page 1047-1049; 2188pp; English.
 CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
 CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
 CC (I) and (II) can have antibacterial activity and therefore can be used
 CC in vaccination. The nucleic acids (I) may be used to produce the
 CC S. epidermidis polypeptides (II) via the production of vectors
 CC containing them which are used to produce hosts cells which express the
 CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
 CC used to vaccinate subjects and to raise antibodies against the bacteria.
 CC The polypeptides may also be used to assay for other inhibitors of their
 CC activity and therefore identify compounds that may be used for the
 CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
 CC AAH5090 represent specifically claimed S. epidermidis genomic DNA
 CC polynucleotide sequences from the present invention. AAH5091 to
 CC AAH5098 represent oligonucleotide sequences and primers which are used
 CC in the exemplification of the present invention.
 CC N.B. The present invention specifically claims all the polynucleotide
 CC sequences given in the sequence listing of the present specification,
 CC however the sequence listing only goes up to SEQ ID NO:4454 so even
 CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
 CC no sequences are present for SEQ ID NO:4455 to 4464.
 XX
 SQ Sequence 4105 BP; 1478 A; 590 C; 744 G; 1293 T; 0 other;

Query Match 18.6%; Score 108.2; DB 22; Length 4105;
 Best Local Similarity 50.7%; Pred. No. 5.7e-21;
 Matches 309; Conservative 0; Mismatches 273; Indels 27; Gaps 1;

QY 1 ATGTTAATGCTTATTGATTATTTTGGCCCTAGTTAGGACGACATTCATCTGGCTTA 60
 DB 1557 ATGATGATCATGTCATGTAATCTGAGTTATCTGATGTGTCATCCCAAGCGGGTA 1498
 QY 61 ATTGGGCGACCTGCCAAGGAATGATATTCGGGACGACGGAAGCGCAACTTAGGC 120
 DB 1497 ATTATTGTAATATTTTAAAAAAGATATAAGACAATACGCTAGTGGAAATACATGGA 1438
 QY 121 GCTACCAATGATCCGTCATGTTGGGTGTAAGAGCTGGTCCGGTCGTCATAGCGGAGAT 180
 DB 1437 GCACTACAGTTTTCGGTGTCTTGGGAAGACCGACGCTGGATTTATGATACGTTTATGAT 1378
 QY 181 ATTTTGAAGGGACACTGCGCACTGCTTTCTCATGTCATGTTGATATTCACCGC 240
 DB 1377 ATTTTCAAGGGATTTATTACAGTCTTTTTCCTACTATGTTCCAGTTCATGCGGATGGT 1318
 QY 241 CTCTCTT-----CGAGGAGTCTTTGGGGTTTTAGGCCAC 273
 DB 1317 GTTATAAGCACCTTCTTTTACAAATGGTTTAAATAGTAGGATTTGTTGCAATACTCGGTCAC 1258

QY 274 GTCTTTCCCATCTTCGCCAAATTTAAAGCGCGTAAAGCGTGGCGACATCAGAGCGGTT 333
 DB 1257 GTGTATCCCAATATATCTGAATTTAATGGCGAAAGACAGTAGTACCAGTGCAGGAGTT 1198
 QY 334 TTGCTATTTTACGACCCCTGTTTATTTATCAGCATGGTTCGGGTATTTCTTCATCTTTTAA 393
 DB 1197 GTATTAGGTGTCATCTTATTTTACTTCTTATCTTGGCAATTATCTTTTCTAGTGATTA 1138
 QY 394 TACTTGACTAAATTTGTTTCTCTCTCATCGATGTTAACAGGGATCTATACGTATATATAT 453
 DB 1137 AAAATCTTTAAATATGTTTCTTTATCAAGTATCATTTGACGAATTTAGTCTGTCGATGGT 1078
 QY 454 AGTTCTTCTTCATGATACGTATTTTATTCATGTTGCTTACCCTGCTCAGTATTTTGTG 513
 DB 1077 TCAATCATCAATCATGATATATTTTACTTCTGCTGTTAGCGGAATTTGTTCAATCATATTA 1018
 QY 514 ATATACAGACACCGAGCGAATTAACAGCAATTAATCAATAAAGACAGACCTAAAGTAAAA 573
 DB 1017 ATRATTCGACACAAATCTAATATATAGTTAGAATTTTAAAGGAGAGAAGCACTAAATATA 958
 QY 574 TGCTTATAA 582
 DB 957 TGCATGTAA 949

RESULT 11
 AAH53731
 ID AAH53731 standard; DNA; 594 BP.
 XX
 AC AAH53731;
 XX
 DT 03-SEP-2001 (first entry)
 XX
 DE S. epidermidis open reading frame nucleotide sequence SEQ ID NO:2855.
 XX Staphylococcus epidermidis SRI strain; infection; diagnosis;
 KW vaccination; endocarditis; ds.
 KW
 XX Staphylococcus epidermidis.
 OS
 XX WO200134809-A2.
 XX
 PD 17-MAY-2001.
 XX
 PF 09-NOV-2000; 2000WO-US30782.
 XX
 PR 09-NOV-1999; 99US-0164258.
 XX
 PA (GLAX) GLAXO GROUP LTD.
 XX
 PI Kimmerly WJ;
 XX
 DR WPI; 2001-316495/33.
 DB P-PSDB; AAG82881.
 PT Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
 PT useful for vaccinating against infections, e.g. endocarditis .
 XX Claim 8; Page 744; 2188pp; English.
 CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
 CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
 CC (I) and (II) can have antibacterial activity and therefore can be used
 CC in vaccination. The nucleic acids (I) may be used to produce the
 CC S. epidermidis polypeptides (II) via the production of vectors
 CC containing them which are used to produce hosts cells which express the
 CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
 CC used to vaccinate subjects and to raise antibodies against the bacteria.
 CC The polypeptides may also be used to assay for other inhibitors of their
 CC activity and therefore identify compounds that may be used for the
 CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
 CC AAH5090 represent specifically claimed S. epidermidis genomic DNA
 CC polynucleotide sequences from the present invention. AAH5091 to

CC AHH5098 represent oligonucleotide sequences and primers which are used
CC in the exemplification of the present invention.
CC N.B. The present invention specifically claims all the polynucleotide
CC sequences given in the sequence listing of the present specification,
CC however the sequence listing only goes up to SEQ ID NO:4454 so even
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
CC no sequences are present for SEQ ID NO:4455 to 4464.
XX
SQ Sequence 594 BP; 181 A; 78 C; 108 G; 227 T; 0 other;

Query Match 18.3%; Score 106.6; DB 22; Length 594;
Best Local Similarity 50.9%; Pred. No. 7.7e-21;
Matches 302; Conservative 0; Mismatches 264; Indels 27; Gaps 1;

QY 17 TGATTATTTTGGCCCTACTGATAGGACGATTCCTGCTGGCTTAATTTGGCGAAGCTG 76

DB 2 TGTAACTCTTGATGATCTGATGCTGATTCCTCCAGCGGGTAAATTTGGTAAATAT 61

QY 77 CCAAGGAATGATATTCGGGAGCAGGAGCGCACTTAGCGGCTACCAATGCAATTC 136

DB 62 TTTTAAAGAAAGATATAAGACAATACGCTAGTGGAAATACTGGAGCACTAACAGTTTC 121

QY 137 GTACATTGGGCTAAAGCTGTTCCGTCGTCATAGCCGGAGATATTTTGAAGGACAC 196

DB 122 GTGTTCTTTGGAAGACCACTGGGATTTATAGTTACCTTTTATAGATATTTTCAAGGATTTA 181

QY 197 TGGCAACTGCATTTCTTCATGTCATGTTGATATTACCCGCTTCTT----- 246

DB 182 TTACAGTCTTTTCCACTAGTGGTCCAGATTCATCGGATGGTGTATTAAGCACTTCT 241

QY 247 -----CGAGGAGCTTTTGGGTTTATAGGCGACCTGTTCCCATCTTCG 289

DB 242 TTACAAATGGTTTAAATAGTAGGATGTTTGGCAATACTCGTCACTATCCAAATATC 301

QY 290 CCAATTTAAAGCGGTAAAGCGTGGCAGATCAGGAGCGGTTGCTATTTTACGCAC 349

DB 302 TGAATTTTAAAGCGGAAAGCAGTAGTACCAGTGCAGGAGTGTATTAAGTGTCAATC 361

QY 350 CCTGTTTATTTATCAGATGTTGGGATTTCTTCATCTTTTATACCTTGTGCTAAATTTG 409

DB 362 CTATTTTACTTCTTCTTGGCAATTAATCTTTTATGCTATTAATAATCTTTAAATATG 421

QY 410 TTTCTCTCTCATCGATGTTAAAGGAGTATACCTGTTATATATAGTTCTTTTGTCCATG 469

DB 422 TTTCTTTATCAAGTATCATTCAGCAATTAGTGTGTGATTTGGTTCATCATCATCATG 481

QY 470 ATACGTATTTATGATTTGCTGTTACCTGCTCCTATTTTCTGATATACAGACCCGAG 529

DB 482 ATTATATTTTACTTGTGCTGTTAGCGGAATGTTTCAATCATATTAATAATTCGACACAAT 541

QY 530 CGAACATTAACGAATTTATCAATAAAGACCACTAAAGTAAATGGTTATTA 582

DB 542 CTAAATATAGTAGAATTTTAAAGGAGAGAACCTAAATTAATTAATGGATGTA 594

RESULT 12

AAS55594

ID AAS55594 standard; DNA; 642 BP.

AC AAS55594;

XX 13-FEB-2002 (first entry)

XX Streptococcus pneumoniae DNA for cellular proliferation protein #165.

XX Antisense; ds; prokaryotic cellular proliferation gene;

KW antibiotic; antibacterial; drug design.

XX Streptococcus pneumoniae.

OS WO2001/0955-A2.

PN 27-SEP-2001.

PD

XX 21-MAR-2001; 2001WO-US09180.

XX 21-MAR-2000; 2000US-191078P.

XX 23-MAY-2000; 2000US-206848P.

XX 26-MAY-2000; 2000US-207272P.

XX 23-OCT-2000; 2000US-242578P.

XX 27-NOV-2000; 2000US-253625P.

XX 22-DEC-2000; 2000US-257931P.

XX 16-FEB-2001; 2001US-269308P.

XX (ELIT-) ELITRA PHARM INC.

XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

XX Yamamoto RT, Xu HH;

XX WPI: 2001-611495/70.

XX P-PSDB: AAU37735.

XX New polynucleotides for the identification and development of
XX antibiotics, comprise sequences of antisense nucleic acids -
XX Claim 27; Seq ID No 9231; 511pp; English.

XX The invention relates to antisense inhibitors of genes essential to
XX prokaryotic cellular proliferation, their use in identifying the
XX genes, their use in the discovery of novel antibiotics, the essential
XX genes themselves and the encoded proteins. The prokaryotes used are
XX Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
XX pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
XX invention is also useful for the identification of potential new targets
XX for antibiotic development. The antisense nucleic acids can also be used
XX to identify proteins used in proliferation, to express these proteins,
XX and to obtain antibodies capable of binding to the expressed proteins.
XX The proteins can be used to screen compounds in rational drug discovery
XX programmes. The antisense nucleic acid sequence is also useful to screen
XX for homologous nucleic acids which are required for cell proliferation in
XX a wide variety of organisms. The present sequence encodes an
XX essential prokaryotic cellular proliferation protein.
XX Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic
XX format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 642 BP; 151 A; 144 C; 122 G; 225 T; 0 other;

XX Query Match 17.5%; Score 101.6; DB 23; Length 642;

XX Best Local Similarity 53.1%; Pred. No. 2.2e-19;

XX Matches 240; Conservative 0; Mismatches 209; Indels 3; Gaps 1;

QY 1 ATGTTAATTTGCTTTATTTGATTATTTTGGCTTCTTATGATAGGAGCATTCCTTCGCTTA 60

DB 1 ATGATTACAATAGTATTTTATTAATCTAGCTATCTGCTGGTTCGATTCATCTCGTCTC 60

QY 61 ATTTGGGCAAGCTTGGCAAGGAATTCATATTCGGGAGCAGCGGCAACTTAGGC 120

DB 61 TGGATTGGCAAGTATTTCTTCAATCAATCTACCGAGCATGTTCTGTAACACATGGA 120

QY 121 GCTACCAATGCAATCCGTACATTTGGTGTAAAAGCTGTTGGTCTCATAGCCGAGAT 180

DB 121 AGCACCACACCTCCCGCATTTTAGTAAAGAGCTGTATGGCAACCTTTGTGATTGAC 180

QY 181 ATTTTGAAGGACACTGGCAACTGCATTTGCCTTTTCTCATGCAATGT---TGATATTCAC 237

DB 181 TTTTCAAAAGGAACCTTAGCAACGCTGCTTCGATTTATTTTTCATCTACAGGCGTTCT 240

QY 238 CGCTTCTTCAGGAGTCTTTTCGGTGTAGCCAGCTGTTTCCCATCTTCGCCAAATTT 297

DB 241 CCTCTCATCTTTGGACTTTTGGCTTTATTCGGCCATACCTTCCTCTATCTTTCAGGATTT 300

QY 298 AAAGCGGTAAAGCGGTGGCGACATCAGGAGCGCTTTTGTCTATTTTACGACCCCTGTGA 357

DB 301 AAAGGTGGTAGGCTGTCCCAACCACTGCTGGAGTGATTTTCGGATTTTGGCGCTATCTTC 360

QY 358 TTTATCAGCATGGTTCGGGTAATCTTCATCTTTTATACCTGACTAAATTTGTTCTCTC 417
 Db 361 TGCTCTACCTTCGGATTAATCTTCCTTGGAGCTCTCTATCTTGGCAGTAATGATTCAC 420
 QY 418 TCATCGATGTTAACAGGATCTATACCTGTTAT 449
 Db 421 TCTAGTGTACAGCATCGATCGCGGCTGTTAT 452

RESULT 13
 AAS55801
 ID AAS55801 standard; DNA; 642 BP.
 XX
 AC AAS55801;
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE Streptococcus pneumoniae DNA for cellular proliferation protein #372.
 XX
 KW Antisense; ds: prokaryotic cellular proliferation gene;
 KW antibiotic; antibacterial; drug design.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN WO200170955-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 21-MAR-2001; 2001WO-US09180.
 XX
 PR 21-MAR-2000; 2000US-191078P.
 PR 23-MAY-2000; 2000US-206848P.
 PR 26-MAY-2000; 2000US-207727P.
 PR 27-OCT-2000; 2000US-242578P.
 PR 27-NOV-2000; 2000US-253625P.
 PR 22-DEC-2000; 2000US-257931P.
 PR 16-FEB-2001; 2001US-269308P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
 PI Yamamoto RT, Xu HH;
 XX
 DR WPI: 2001-611495/70.
 DR P-PSDB; AAU37942.
 XX
 PT New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids -
 XX
 PS Claim 27; Seq ID No 9438; 511pp; English.

CC The invention relates to antisense inhibitors of genes essential to
 CC prokaryotic cellular proliferation, their use in identifying the
 CC genes, their use in the discovery of novel antibiotics, the essential
 CC genes themselves and the encoded proteins. The prokaryotes used are
 CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
 CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
 CC invention is also useful for the identification of potential new targets
 CC for antibiotic development. The antisense nucleic acids can also be used
 CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence encodes an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 642 BP; 151 A; 144 C; 122 G; 225 T; 0 other;

Query Match 17.5%; Score 101.6; DB 23; Length 642;
 Best Local Similarity 53.1%; Pred. No. 2.2e-19;
 Matches 240; Conservative 0; Mismatches 209; Indels 3; Gaps 1;
 QY 1 ATGTTAATGCTTTATGATTATTTGGCTACTTATGATAGGCAGCATCTCCATCTGCTCTC 60
 Db 1 ATGATTACANATAGTTTTTATTAATCTAGCCTATCTGCTGGTTCGATTCCATCTGCTCTC 60
 QY 61 ATTGTGGCAAGCTTGGCCAAAGGAATGATATTCGGGAGCAGCGAAGCGCAACTTAGGC 120
 Db 61 TGGATTGGACAAGTATTTCTTCAAAATCAATCTACCGAGCATGGTTCTGGTAACACTGGA 120
 QY 121 GCTACCAATGCAATTCCTGATACATTGGGTGTAAGAGCTGGTTCGTCGTCATAGCCGAGAT 180
 Db 121 ACACCAACACCTTCCGCATTTTAGGTAAAGAAAGCTGGTATGGCAACCTTTGCTGATTGAC 180
 QY 181 ATTTTGAAGGACACTGGCAACTGCATTGCTTTTCTCATGCACTGT--TGATATTTCAC 237
 Db 181 TTTTTCAAAGGACCTAGCAACGCTGCTCGGATTATTTTTCATCTACAAGGCGCTTTCT 240
 QY 238 CGGCTTCTTGGCAGGAGTCTTTTGGGTTTATAGGCCACCGTGTTCCTCCATCTCGCCAAATTT 297
 Db 241 CCTCTCATCTTTGGACTTTTGGCTGTTATCGCCATACCTTCCCTATCTTTGCGAGGATTT 300
 QY 298 AAGGCGGTAAGCGCTGGCGACATCAGGAGCGGTTTTCGTTATTTAGCACCCTCTGTTA 357
 Db 301 AAGGTGGTAAGGCTGTGCAACCAAGCTGCTGGAGTGATTTTCGGATTTCGCCCTATCTTC 360
 QY 358 TTTATCAGCATGGTTCGGGTATCTTCTCATCTTTTATATACCTGACTAAATTTGTTCTCTC 417
 Db 361 TGCTCTACCTTGGGATTAATCTCTTTTGGAGCTCTCTATCTTGGCAGTATGATTTCAC 420
 QY 418 TCATCGATGTTAACAGGATCTATACCTGTTAT 449
 Db 421 TCTAGTGTACAGCATCGATCGCGGCTGTTAT 452

RESULT 14
 AAA09181
 ID AAA09181 standard; DNA; 642 BP.
 XX
 AC AAA09181;
 XX
 DT 10-AUG-2000 (first entry)
 XX
 DE S. pneumoniae S-ynes coding sequence.
 XX
 KW S-ynes; survival; antibacterial; inhibitor; ds.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN WO200020627-A1.
 XX
 PD 13-APR-2000.
 XX
 PF 30-SEP-1999; 99WO-US222665.
 XX
 PR 30-SEP-1998; 98US-0163445.
 XX
 PA (MILL-) MILLENNIUM PHARM INC.
 XX
 PI Fritz C, Youngman P, Guzman L;
 XX
 DR WPI: 2000-303799/26.
 DR P-PSDB; AAY92246.
 XX
 PT Methods for identifying an antibacterial agent for treating
 PT Streptococcus pneumoniae infections comprises detecting an interaction
 PT between a ynes polypeptide and a test compound
 XX
 PS Disclosure; Fig 1; 65pp; English.

Streptococcus pneumoniae ynes gene encodes a polypeptide (S-ynes)
essential for survival for a wide range of bacteria. Identifying an
antibacterial agent comprises contacting a ynes polypeptide (S-ynes) with
a test compound and detecting an interaction of the test compound with
the S-ynes polypeptide which indicates that the compound is an
antibacterial agent. Alternatively, detecting a decrease in function of
the polypeptide contacted with the test compound and determining whether
the compound inhibits growth of bacteria, relative to the growth of
bacteria cultured in the absence of a test compound where inhibition of
growth indicates the compound is an antibacterial agent. Inhibitors of
S-ynes function are useful for treating a Streptococcus pneumoniae
infection in mammals.

Sequence 642 BP; 151 A; 143 C; 122 G; 226 T; 0 other;

Query Match 17.2%; Score 100; DB 21; Length 642;

Best Local Similarity 52.9%; Pred. No. 6.5e-19;

Matches 239; Conservative 0; Mismatches 210; Indels 3; Gaps 1;

QY 1 ATGTTAATGCTTTATTCATTAATTTGGCTTACTTGTATAGGCAGCATTCCTCCATCTGGCTTA 60
DB 1 ATGATTACAATAGTTTATTAATCTAGCCATCTGCTGGTTGATTCATCTGGCTTC 60
QY 61 ATTGGGCAAGCTTGGCAAGGAATTTGATATTCGGGAGCACGGAAGCGCAACTTAGGC 120
DB 61 TGGATTGGACAAGTATTTCTTCAATCAATCTACGCGAGCATGGTTCTGGTAACACTGGA 120
QY 121 GCTACCAATGATTCGATACATTTGGGTAAAGCTGGTTGCGTGCATAGCGGAGAT 180
DB 121 ACGACCAACACCTTCCGCAATTTAGGTAAAGAAAGCTGGTATGGCAACCTTTGTGATGAC 180
QY 181 ATTTTAAAGGACACTGGCAACTGCATTGCTTTTCTCATGCACTGTTTGTATTTTAC 237
DB 181 TTTTCAAGGACACTGGCAACTGCATTGCTTTTCTCATGCACTGTTTGTATTTTAC 240
QY 238 CGCTCTTTCGAGGAGTCTTTGGGTTTAAAGCCACGCTTTTCCCATCTTCCGCCAATTT 297
DB 241 CCTCTCATCTTTGGGACTTTTGGCTGTTATTCGGCCATACCTTCCCTATCTTTCGAGGATTT 300
QY 298 AAAGCGGTAAAGCGGTGGGACATCAGGAGCGGTTTGTCTATTTTACGACCCCTGTTA 357
DB 301 AAAGGTGGTAAGGCTGTGCAACACAGTGTGAGGTGATTTTGGGATTTTGGCGCTATCTTC 360
QY 358 TTTATCAGCATGTTTGGGCTATTTCTTCTATCTTTTATATCTTACCTAAATTTTGTCTCTC 417
DB 361 TGTCTCTACCTTGGCATATCTCTTTGGAGCTCTCTATCTTGGCAGTATGATTTTCACTG 420
QY 418 TCATCAGATGTTTAAAGGATCTATATCTTTAT 449
DB 421 TCTAGTGTACAGCATCGATTGCGGCTGTTAT 452

RESULT 15

AAV52170

ID AAV52170 standard; DNA; 6171 BP.

XX

AC AAV521170;

XX

DT 23-OCT-1998 (first entry)

XX

DE Streptococcus pneumoniae genome fragment SEQ ID NO:37.

XX

KW Streptococcus pneumoniae; S. pneumoniae; genome: diagnosis; assay;

XX

KW computer readable medium; vaccine; pharmaceutical composition; ds.

XX

OS Streptococcus pneumoniae.

XX

PN W09818931-A2.

XX

PD 07-MAY-1998.

XX

PF 30-OCT-1997; 97WO-US19588.

XX

PR 31-OCT-1996; 96US-0029960.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX

PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;

PI Kunsch CA, Rosen CA;

XX

DR MPI; 1998-272225/24.

XX

PT Computer-readable medium with recorded Streptococcus pneumoniae
polynucleotide sequences - useful in diagnostic kits and assays, and
pharmaceutical compositions and vaccines for Streptococcus
pneumoniae

XX

PS Claim 1; Page 371-374; 1409pp; English.

XX

CC The present invention describes a computer readable medium which has
the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524)
recorded on it, or a representative fragment or a sequence at least 95%
identical to SEQ ID NO:1 to 391. The nucleotide sequences depicted in
SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from
Streptococcus pneumoniae. The present invention also describes an
isolated nucleic acid molecule encoding a homologue of any of the
fragments of the S.pneumoniae genome (SEQ ID NO:1 to 391) where the
nucleic acid molecule is produced by a process comprising: (a) screening
a genomic DNA library using as a probe a target sequence defined by any
of the sequences in SEQ ID NO:1 to 391, identifying members of the
library which contain sequences that hybridize to the target sequence and
isolating the nucleic acid molecules from the members; or (b) isolating
mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid
molecules whose nucleotide sequence is homologous to amplification
primers derived from the fragment of the S. pneumoniae genome to prime
the amplification and isolating the amplified sequences. The computer
readable medium can be used in a computer-based system for identifying
fragments of the S. pneumoniae genome of commercial importance, or
expression modulating fragments of the S. pneumoniae genome. Products
from the present invention can be used in diagnosis kits and assays, and
pharmaceutical compositions and vaccines for S. pneumoniae.

XX Sequence 6171 BP; 1645 A; 1424 C; 1188 G; 1914 T; 0 other;

Query Match 17.2%; Score 100; DB 19; Length 6171;

Best Local Similarity 52.9%; Pred. No. 1.6e-18;

Matches 239; Conservative 0; Mismatches 210; Indels 3; Gaps 1;

QY 1 ATGTTAATGCTTTATTCATTAATTTGGCTTACTTGTATAGGCAGCATTCCTCCATCTGGCTTA 60
DB 5149 ATGATTACAATAGTTTATTAATCTAGCCATCTGCTGGTTGATTCCTCCATCTGGCTTC 5208
QY 61 ATTGTGGGCAAGCTTGGCAAGGAATTTGATATTCGGGAGCACGGAAGCGCAACTTAGGC 120
DB 5209 TGGATTGGACAAGTATTTCTTCAATCAATCTACGCGAGCATGGTTCTGGTAACACTGGA 5268
QY 121 GCTACCAATGATTCGCTACATTTGGGTAAAGCTGGTTGCGTGCATAGCGGAGAT 180
DB 5269 ACGACCAACACCTTCCGCAATTTAGGTAAAGAAAGCTGGTATGGCAACCTTTGTGATTCAC 5328
QY 181 ATTTTGAAGGACACTGGCAACTGCATTGCTTTTCTCATGCACTGTTTGTATTTTAC 237
DB 5329 TTTTTCAAAGGAACCTTAGCAACGCTGCTTCCGATTTATTTTCAATCAAGGCTTTCT 5388
QY 238 CGCTTCTTGGAGGAGTCTTTTGGGTTTATAGGCCACGCTGTTTCCCATCTTCCGCAAAATTT 297
DB 5389 CCTCTCATCTTTGGACTTTTGGCTGTTATCGCCATACCTTCCCTATCTTTGCGAGGATTT 5448
QY 298 AAAGCGGTAAAGCGGTGGGACATCAGGAGCGGTTTGTGCTATTTTACGCAACCCCTGTTA 357
DB 5449 AAAGGTGGTGAAGCTGTGCAACACAGTGTCTGGAGTATTTTTCGGATTTTTCGCTATCTTC 5508
QY 358 TTTATCAGATGTTTGGCGTATTTCTTCTATCTTTTATATCTTATGACTAAATTTGTTCTCTC 417
DB 5509 TGTCTCTACCTTGGGATTTATCTTTTGGAGCTCTCTATCTTGGCAGATGATGATTTCACTG 5568

5

Search completed: December 24, 2002, 22:48:36
Job time : 459.703 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 24, 2002., 22:16:24 ; Search time 3199.81 Seconds
(without alignments)
2945.728 Million cell updates/sec

Title: US-10-068-080-4
Perfect score: 582
Sequence: 1 atgttaatgctttattgat.....ctaaagtaaaatgggtataa 582

Scoring table: IDENTITY_NUC
Gapox 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST.*
1: em_estba.*
2: em_esthum.*
3: em_estin.*
4: em_estnu.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_hic.*
9: gb_estil.*
10: gb_est2.*
11: gb_hic.*
12: gb_est3.*
13: gb_est4.*
14: gb_est5.*
15: em_estfun.*
16: em_estom.*
17: gb_gss.*
18: em_gss_hum.*
19: em_gss_inv.*
20: em_gss_pln.*
21: em_gss_vrt.*
22: em_gss_fun.*
23: em_gss_mam.*
24: em_gss_mus.*
25: em_gss_other.*
26: em_gss_pro.*
27: em_gss_rod.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
C 1	49.2	8.5	330	9	AL513817
C 2	49.2	8.5	389	9	AL513693
C 3	48.4	8.3	508	14	BQ391243
4	47.8	8.2	1101	17	QNSOI7NP
5	47.4	8.1	740	17	AQ331048
C 6	47.2	8.1	329	9	AL513719

7	46.4	8.0	458	9	AL514085
C 8	46.2	7.9	477	14	BQ390789
9	45.6	7.8	379	9	AL514359
C 10	45.2	7.8	514	14	BQ521704
11	44.6	7.7	413	17	B2057A
C 12	44.4	7.6	329	9	AL514627
C 13	44.4	7.6	335	9	AL513597
14	44.4	7.6	414	17	A2524410
15	44.2	7.6	413	14	BQ076552
16	44.2	7.6	414	14	BQ077621
C 17	44.2	7.6	415	10	AV657079
18	44.2	7.6	474	14	BQ076863
19	44	7.6	365	9	AL515373
20	44	7.6	900	17	A2682023
21	44	7.6	1270	13	BM456675
C 22	43.8	7.5	630	17	AG154621
23	43.8	7.5	743	13	BJ446581
24	43.6	7.5	360	14	BQ400218
C 25	43.6	7.5	533	14	BQ396711
C 26	43.6	7.5	761	17	AG152484
C 27	43.4	7.5	638	9	AL513901
C 28	43.2	7.4	454	9	AL513951
C 29	43.2	7.4	651	17	CNS03PNO
C 30	43.2	7.4	867	17	CNS00CX5
C 31	43	7.4	313	14	BQ391420
C 32	43	7.4	338	14	C90212
C 33	43	7.4	341	17	CNS00606
34	42.8	7.4	300	17	CNS0076L
C 35	42.3	7.4	392	9	AL514511
C 36	42.3	7.4	423	9	AL513863
C 37	42.3	7.4	889	17	BH134799
C 38	42.3	7.4	899	17	BH154279
39	42.8	7.4	957	17	BH130614
C 40	42.8	7.4	1045	17	CNS03YE4
C 41	42.6	7.3	367	14	BQ525045
C 42	42.6	7.3	870	17	CNS06022
C 43	42.6	7.3	924	17	CNS01GHN
C 44	42.4	7.3	1074	17	CNS01082
C 45	42.2	7.3	637	17	CNS0245B

ALIGNMENTS

RESULT 1
AL513817/c
LOCUS AL513817 LTI_NFL006_PL2 Homo sapiens cDNA clone CLOBA0092E06 330 bp mRNA linear EST 13-FEB-2001
DEFINITION prime, mRNA sequence.
ACCESSION AL513817
VERSION AL513817.1 GI:12777311
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 330)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
Bp 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
Location/Qualifiers
1..330
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CLOBA0092E06"
/clone_lib="LTI_NFL006_PL2"
/tissue_type="placenta"
/note="Vector: pCMVSPORT 6; Site 1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end

enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 183 a 28 c 17 g 88 t 14 others
ORIGIN

Query Match 8.5%; Score 49.2; DB 9; Length 330;
Best Local Similarity 47.1%; Pred. No. 0.054;

Matches 114; Conservative 10; Mismatches 118; Indels 0; Gaps 0;

OY 332 TTTTGGCTATTTACGACCCCTGTTATTTATCATCGATGGTTGCGGTATCTTCATCTTTT 391

||||| ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||

Db 277 TTTTATTTTTCYCCYTT 218

||||| ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||

OY 392 TATACCTGACATAATTTGTTCTCTCATCGATGTTAAACAGGATCATACCTTATAT 451

||||| ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||

Db 217 TTTTCYTT 158

||||| ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||

OY 452 ATAGTTTCTTGTCCATGATACGTATTTATGATGTCGTACCCCTGCTCACTATTTTG 511

||||| ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||

Db 157 TTTTATTTTTCYCCYTT 98

||||| ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||

OY 512 TGATATACAGACCGGACGAACTTAACGAAATTAATCAATAAAGACCTAAAGTAA 571

||||| ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||

Db 97 TTAAAAAATTTTAAAAAATTTTAAAAAATTTTAAAAAATTTTAAAAAATTTTAAAA 38

||||| ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||

OY 572 AA 573

||

Db 37 AA 36

RESULT 2
AL513693 389 bp mRNA linear EST 13-FEB-2001
LOCUS
DEFINITION AL513693 LTI_NFL006_PL2 Homo sapiens cDNA clone CLOBA007D12 3 prime, mRNA sequence.

ACCESSION AL513693
VERSION AL513693.1 GI:12777187
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 389)
Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
Full-length cDNA libraries and normalization

TITLE Unpublished (2001)
JOURNAL
COMMENT Contact: Genoscope

Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
Location/Qualifiers
1. 389

source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CLOBA007D12"
/clone_lib="LTI_NFL006_PL2"

/tissue_type="placenta"
/note="vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end

enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by

Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371

Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 201 a 13 c 5 g 141 t 29 others
ORIGIN

Query Match 8.5%; Score 49.2; DB 9; Length 389;
Best Local Similarity 45.9%; Pred. No. 0.053;

Matches 111; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

OY 332 TTTTGGCTATTTACGACCCCTGTTATTTATCATCGATGGTTGCGGTATCTTCATCTTTT 391

||||| ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||

Db 327 TTTTATTTTTCYCCYTT 268

||||| ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||

OY 392 TATACCTGACATAATTTGTTCTCTCATCGATGTTAAACAGGATCATACCTTATAT 451

||||| ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||

Db 267 TTTTATTTTTCYCCYTT 208

||||| ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||

OY 452 ATAGTTTCTTGTCCATGATACGTATTTATGATGTCGTACCCCTGCTCACTATTTTG 511

||||| ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||

Db 207 TTTTATTTTTCYCCYTT 148

||||| ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||

OY 512 TGATATACAGACCGGACGAACTTAACGAAATTAATCAATAAAGACCTAAAGTAA 571

||||| ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||

Db 147 TTAAAAAATTTTAAAAAATTTTAAAAAATTTTAAAAAATTTTAAAAAATTTTAAAA 88

||||| ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : |||||

OY 572 AA 573

||

Db 87 AA 86

RESULT 3

BO391243/c

LOCUS

DEFINITION BO391243

ACCESSION BO391243

VERSION BO391243.1

KEYWORDS EST.

SOURCE

ORGANISM

western clawed frog.

Silurana tropicalis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Anphibia; Batrachia; Anura; Mesobatrachia; Pipidae;

Xenopodinae; Silurana.

REFERENCE 1 (bases 1 to 508)

NIH-XCG http://image.llnl.gov/image/html/xenopuslib_info.shtml.

AUTHORS

National Institute of Child Health and Human Development, National

Cancer Institute, Xenopus Gene Collection

Unpublished (2002)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgabbs@mail.nih.gov

CDNA Library Preparation:

CDNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL

DNA Sequencing by: National Institutes of Health Intramural

Sequencing Center (NISC)

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

info@image.llnl.gov

Plate: LLAM11782 row: K column: 5

Seq primer: -21M13 forward primer (ABI).

Location/Qualifiers

1. 508

source

/organism="Silurana tropicalis"

/db_xref="taxon:8364"

/clone="IMAGE:5308924"

/clone_lib="NICHD XGC Emb5"

/tissue_type="gastrula"

/dev_stage="embryo, stages 10-13"

/lab_host="DH10B (phage-resistant)"

/note="vector: pCMV-SPORT6.cdb; Site_1: NotI; Site_2:

EcoRV; Cloned unidirectionally. Primer: Oligo dr. Average

insert size 2.0 kb. Constructed by Invitrogen. Note: This

is a Xenopus Gene Collection (XGC) library."

BASE COUNT 316 a 42 c 56 g 94 t

ORIGIN

```

Query Match      8.38; Score 48.4; DB 14; Length 508;
Best Local Similarity 50.08; Pred. No. 0.081;
Matches 121; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

QY 332 TTTTGCATATTTAGCACCCCTGTTATTTATCATCAGATGGTTCGGGTATCTTCTATCTTTT 391
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 294 TTTTGTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 235
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 392 TATCTTGACATAAATTTCTCTCTCATCGATGTTAAACAGGATCTATACCTGTTATAT 451
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 234 TTTTGTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 175
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 452 ATAGTTTCTTTGTCATGATACGATTTATTTGATTTGCTTTACCTGCTCACTATTTTGT 511
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 174 TTTTGTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 115
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 512 TGATATACAGACACCGAGCGAACATTAACAGCAATTAACATAAACAAGCAAGCAAGTAA 571
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 114 TTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 55
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 572 AA 573
      ||
DB 54 AA 53

RESULT 4
CNS017NP
LOCUS
DEFINITION Drosophila melanogaster genome survey sequence 17 end of BAC
            BACN37H08 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION AL108271
VERSION AL108271.1 GI:5628575
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS
TITLE Direct Submission
JOURNAL
COMMENT Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaut at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBelOBAC11.
FEATURES
            source
            1. .1101
               Location/Qualifiers
               /organism="Drosophila melanogaster"
               /db_xref="taxon:7227"
               /clone="BACN37H08"
               /clone_lib="drosBAC"
               /plasmid="pBelOBAC11"
               /note="end : T7"
BASE COUNT 281 a 179 c 128 g 306 t 207 others
ORIGIN

Query Match      8.28; Score 47.8; DB 17; Length 1101;
Best Local Similarity 37.19; Pred. No. 0.1;
Matches 96; Conservative 45; Mismatches 118; Indels 0; Gaps 0;

QY 324 AGGAGCGTTTGTCTATTTACGACCCCTGTTATTTATCATCAGATGGTTCGGGTATCTTT 383
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 637 ACGBKCAAACTTAATTTTCTTCTTCCTTCCTTCGTCATCAAACTTCATCCCAATTTT 696
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 384 CATCTTTTATACCTGACTAAATTTGTTCTCTCTCATCGATGTTAAACAGGATCTATAC 443
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 697 CAMCTATGTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 756
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 444 TGTATATATATAGTTTCTTTTCCATGATACGATGATTTATTTATTTATTTATTTATTTAT 503
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 757 YCTAAAAAAGCGTTTATTTTATTTTAAATTTAAATAAATAAATAAATAAATAAATAAATA 816
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 504 TATTTTGTGTATACAGACACCGACGCAATTAACAGCAATTTATCAATAAACAAGCAACC 563
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 817 KTCCTTTTATTTATTAHCHAHATARSATATAAATAAATAAATAAATAAATAAATAAATA 876
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 564 TAAAGTAAATGGTTATAA 582
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 877 GCGDAWRATTTTAAATAA 895
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 5
AQ331048
LOCUS
DEFINITION nxb0048L01f CUGI Rice BAC Library Oryza sativa genomic clone
            nxb0048L01f, DNA sequence.
ACCESSION AQ331048
VERSION AQ331048.1 GI:4122898
KEYWORDS GSS.
SOURCE Cryza sativa.
ORGANISM Cryza sativa.
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 740)
AUTHORS Wing, R.A. and Dean, R.A.
TITLE A BAC End Sequencing Framework to Sequence the Rice Genome
JOURNAL Unpublished (1998)
COMMENT Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: TAATACGACTCACTATAGGG
            Class: BAC ends
            High quality sequence stop: 352.
FEATURES
            source
            1. .740
               Location/Qualifiers
               /organism="Oryza sativa"
               /strain="Japonica"
               /cultivar="Nipponbare"
               /db_xref="taxon:4530"
               /clone="nxb0048L01f"
               /clone_lib="CUGI Rice BAC Library"
               /tissue_type="Leaf"
               /lab_host="E. coli DH10B"
               /note="Vector: pBelOBAC11; Site:1: HindIII; Site:2:
               HindIII; Rice is one of two most popular grains in the
               world. Half of the world population especially those
               inhabiting highly populated areas of the humid tropics
               and subtropics, rely on rice as their primary source of
               carbohydrate. Monocotyledonous rice is a diploid plant
               (2n=24) with a haploid genome equivalent of 431 Mbp
               (Arumuganathan and Earle, 1991). The relatively small
               genome of rice, three times larger than that of
               Arabidopsis, makes it suitable for genomic studies. In
               order to facilitate positional cloning, physical mapping
               and genome sequencing of rice, we have constructed a BAC
               library from Oryza sativa, Nipponbare variety. The
               library contains 36,864 clones with an average insert size
               of 128.5 Kb providing 10.9 haploid genome equivalents. The
               deep coverage allows the isolation a particular sequence
               with a probability of 99.9 %. Two high density filters,
               each containing 18,432 clones (doubly spotted), represent
               the whole library for colony screening."
BASE COUNT 101 a 62 c 90 g 436 t 51 others

```

Query Match	8.18;	Score 47.2;	DB 9;	Length 329;
Best Local Similarity	37.68;	Pred. No. 0.17;		

50 89
392 TATACTTGACATAAATTTGTTTCTCTCTCATCGATGTTAACAGGGATCTATACTGTTATAT 451
0Y :

[illegible]

```

RESULT 10
B0521704/c
LOCUS      B0521704      514 bp      mRNA      linear      EST 10-JUN-2002
DEFINITION NISC_nll2ell.x1 NICHDXGC_Emb7 Silurana tropicalis cDNA clone
ACCESSION  B0521704
KEYWORDS   IMAGE:5336181 3', mRNA sequence.
SOURCE     B0521704.1 GI:21380583
           EST.
           western clawed frog.
ORGANISM   Silurana tropicalis
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
           Xenopodinae; Silurana.
REFERENCE  1 (bases 1 to 514)
AUTHORS   NIH-XGC http://image.llnl.gov/image/html/xenopuslib_info.shtml.
TITLE     National Institute of Child Health and Human Development, National
           Cancer Institute, Xenopus Gene Collection
JOURNAL   Unpublished (2002)
COMMENT   Contact: Robert Strausberg, Ph.D.
           Email: cgapbs-r@mail.nih.gov
           cDNA Library Preparation:
           cDNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL
           DNA Sequencing by: National Institutes of Health Intramural
           Sequencing Center (NISC)
           Clone distribution: NCI-CGAP clone distribution information can be
           found through the I.M.A.G.E. Consortium/LLNL at:
           info@image.llnl.gov
           Plate: LHAM11853 row: J column: 22
           Seq primer: -21M13 forward primer (ABI).
FEATURES   source
           1..514
           /organism="Silurana tropicalis"
           /db_xref="taxon:8364"
           /clone="IMAGE:5336181"
           /clone_lib="NICHDXGC_Emb7"
           /tissue_type="tailbud"
           /dev_stage="embryo, stages 20-27"
           /lab_host="DH10B (phage-resistant)"
           /note="Vector: pCMV-SPORT6.cdb; Site_1: NotI; Site_2:
           EcoRV; Cloned unidirectionally. Primer: Oligo dt. Average
           insert size 2.1 Kb. Constructed by Invitrogen. Note: This
           is a xenopus Gene Collection (XGC) library."
BASE COUNT 341 a      9 c      5 g      159 t
ORIGIN
Query Match      7.8%; Score 45.2; DB 14; Length 514;
Best Local Similarity 49.2%; Pred. No. 0.51;
Matches 119; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

Qy 332 TTTTCGCTATTTTACGCACCCCGTGTATTATATACAGATGGTTGCGGTATTCATCTTTT 391
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 335 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 276
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 392 TATACCTGACTAAATTTGTTCTCTCTCATCGATGTTAACAGGGATCTATCTGTATAT 451
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 275 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 216
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 452 ATAGTTTCTTGTCCATGATAGTATTTATGATTTGTCGTACCGTCTCACTATTTTG 511
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 215 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 156
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 512 TCATATACAGACACCGAGCAATTAACGAATTTATCAATAAACAACCACTAAAGTAA 571
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 155 TTTTATATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 96
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 572 AA 573
      ||
Db 95 AA 94

RESULT 11
B20574
LOCUS      B20574

```

```

DEFINITION T20L8-T7 TAMU Arabidopsis thaliana genomic clone T20L8, DNA
sequence.
ACCESSION  B20574
VERSION    B20574.1 GI:2395628
KEYWORDS   GSS.
SOURCE     Arabidopsis thaliana
           thale cress.
           Arabidopsis thaliana
           Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
           Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
           Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE  1 (bases 1 to 413)
AUTHORS   Peng,J., Dewar,K., Buehler,E., Kim,C., Li,Y., Shinn,P., Sun,H. and
           Ecker,J.
TITLE     BAC End Sequences at ATGC
JOURNAL    Unpublished (1997)
COMMENT    Other_GSSs: T20L8-Sp6
           Contact: Ecker J.
           Arabidopsis Thaliana Genome Center
           University of Pennsylvania
           Dept. of Biology, University of Pennsylvania, Philadelphia, PA
           19104
           Tel: 215-898-9384
           Fax: 215-898-8780
           Email: jecker@atgenome.bio.upenn.edu
           Seq primer: T7
           Class: BAC ends
           High quality sequence start: 95
           High quality sequence stop: 103.
FEATURES   source
           1..413
           /organism="Arabidopsis thaliana"
           /strain="Columbia"
           /db_xref="taxon:3702"
           /clone="T20L8"
           /clone_lib="TAMU"
           /sex="hermaphrodite"
           /note="Vector: BelOBACII; Site_1: HindIII; Site_2: HindIII
           ; Produced by Rod Wing"
BASE COUNT 20 a      68 c      24 g      293 t      8 others
ORIGIN
Query Match      7.7%; Score 44.6; DB 17; Length 413;
Best Local Similarity 53.1%; Pred. No. 0.75;
Matches 95; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

Qy 332 TTTTGCATTTTACGCACCCCGTGTATTATATACAGATGGTTGCGGTATTCATCTTTT 391
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 234 TTTTCTTTTTCGCCCGCTTTTCTTTTTCGCCCGCTTTTCTTTTTCCTTTTCTTTTCT 293
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 392 TATACCTGACTAAATTTGTTCTCTCTCATCGATGTTAACAGGGATCTATCTATATAT 451
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 294 TTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 353
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 452 ATAGTTTCTTGTCCATGATAGTATTTATGATTTGTCGTACCGTCTCACTATTTT 510
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 354 CTCTCTCTTTTCCCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 412
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

RESULT 12
AL514627/c
LOCUS      AL514627      329 bp      mRNA      linear      EST 13-FEB-2001
DEFINITION AL514627 LTL_NFL006_PL2 Homo sapiens cDNA clone CLOBB002E11 3
prime mRNA sequence.
ACCESSION  AL514627
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 329)
AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE     Full-length cDNA libraries and normalization

```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 24, 2002, 22:25:54 ; Search time 84.875 Seconds
(without alignments)
2102.926 Million cell updates/sec

Title: US-10-068-080-4

Perfect score: 582
Sequence: 1 atgttaattgctttattgat.....ctaaagataaatggtttataa 582

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PTUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	582	100.0	582	4	US-09-222-938A-71
2	108.2	18.6	615	4	US-09-134-001C-2402
3	100	17.2	6171	4	US-08-961-527-37
4	76.6	13.2	344	4	US-09-222-938A-14
5	48.8	8.4	7218	1	US-08-232-463-14
6	37	6.4	19124	2	US-08-487-826B-13
7	35.8	6.2	1723	4	US-09-181-336-14
8	34.2	5.9	1320	1	US-08-257-073-15
9	34.2	5.9	1482	4	US-08-098-327E-41
10	34.2	5.9	1482	4	US-08-462-625-41
11	34.2	5.9	1493	4	US-08-098-327E-38
12	34.2	5.9	1493	4	US-08-462-625-38
13	34	5.8	950	4	US-08-098-327E-32
14	34	5.8	950	4	US-08-462-625-32
15	33.8	5.8	152331	3	US-09-128-155-16
16	33.8	5.8	176373	3	US-09-128-155-17
17	33.4	5.7	375	4	US-09-134-001C-2507
18	33	5.7	3275	4	US-09-370-838-151
19	32.8	5.6	473	1	US-08-764-100-16
20	32.8	5.6	4970	1	US-08-764-100-14
21	32.8	5.6	4970	1	US-08-764-100-20
22	32.6	5.6	1482	4	US-08-098-327E-45
23	32.6	5.6	1482	4	US-08-462-625-45
24	32.4	5.6	15611	4	US-09-441-340-1
25	32	5.5	994	4	US-09-018-584A-42
26	32	5.5	2178	4	US-09-334-601-6
27	32	5.5	2288	4	US-09-334-601-1

28 32 5.5 2448 1 US-08-526-964-2 Sequence 2, Appl
29 32 5.5 2448 2 US-08-946-617-2 Sequence 2, Appl
30 32 5.5 2448 3 US-09-031-897-2 Sequence 2, Appl
31 32 5.5 5203 4 US-09-257-770-1 Sequence 1, Appl
32 31.8 5.5 262 2 US-08-867-030B-3 Sequence 3, Appl
33 31.8 5.5 262 5 PCT-US95-06119-3 Sequence 3, Appl
34 31.4 5.4 318 4 US-09-134-001C-2112 Sequence 2112, Ap
35 31.4 5.4 1956 4 US-08-559-896B-1 Sequence 1, Appl
36 31.4 5.4 2187 4 US-09-134-001C-2131 Sequence 2131, Ap
37 31.4 5.4 2293 4 US-09-336-643A-5 Sequence 5, Appl
38 31.2 5.4 1338 4 US-08-887-534A-60 Sequence 60, Appl
39 31.2 5.4 3494 4 US-09-334-601-5 Sequence 5, Appl
40 31.2 5.4 6124 4 US-08-213-419B-3 Sequence 3, Appl
41 31 5.3 427 4 US-09-397-787-209 Sequence 209, App
42 31 5.3 2013 4 US-09-134-001C-103 Sequence 103, App
43 31 5.3 2447 2 US-09-014-969-14 Sequence 14, Appl
44 31 5.3 2984 4 US-09-605-785-335 Sequence 335, App
45 31 5.3 2984 4 US-09-439-313-335 Sequence 335, App

ALIGNMENTS

RESULT 1
US-09-222-938A-71
; Sequence 71, Application US/09222938A
; Patent No. 5437108
; GENERAL INFORMATION:
; APPLICANT: Youngman, Philip
; APPLICANT: Fritz, Christian
; APPLICANT: Murphy, Christopher
; APPLICANT: Guzman, Luz-Maria
; TITLE OF INVENTION: ESSENTIAL BACTERIAL GENES AND THEIR USE
; FILE REFERENCE: 07334/060001
; CURRENT APPLICATION NUMBER: US/09/222,938A
; CURRENT FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 71:
; LENGTH: 582
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(579)
US-09-222-938A-71

Query Match 100.0%; Score 582; DB 4; Length 582;
Best Local Similarity 100.0%; Pred. No. 1e-169;
Matches 582; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGTTAATGCTTTATTGATTATTTGGCTACTTGTATAGGACGAGCATTCATCTGGCTTA 60
|||||
DB 1 ATGTTAATGCTTTATTGATTATTTGGCTACTTGTATAGGACGAGCATTCATCTGGCTTA 60
QY 61 ATGTTGGGCAAGCTTCCCAAGGAATTGATTTGGGAGCACCAGCGCAACTTAGGC 120
|||||
DB 61 ATGTTGGGCAAGCTTCCCAAGGAATTGATTTGGGAGCACCAGCGCAACTTAGGC 120
QY 121 GCTACCAATGCAATCCGTCATATTTGGGTGTAAGAGCTGGTTCGTCATAGCGCGGAGAT 180
|||||
DB 121 GCTACCAATGCAATCCGTCATATTTGGGTGTAAGAGCTGGTTCGTCATAGCGCGGAGAT 180
QY 181 ATTTTGAAGGGACACTGGCACTGATGCGCTTTTCTCATGATGTTGATATTCACCCG 240
|||||
DB 181 ATTTTGAAGGGACACTGGCACTGATGCGCTTTTCTCATGATGTTGATATTCACCCG 240
QY 241 CTTCTTGCAGAGCTCTTTGGGGTTTGTAGCCACGCTGTTCCCATCTTCGCCAAATTTAA 300
|||||
DB 241 CTTCTTGCAGAGCTCTTTGGGGTTTGTAGCCACGCTGTTCCCATCTTCGCCAAATTTAA 300
QY 301 GCGGTAAAGCCCTGGCGACATCAGGAGCGTTTTCGTATTTTACGACCCCTGTTATT 360
|||||

Db 301 GCGGTAAGCCGTGGCGACATCAGGAGCGGTTTTCGTATTTTACGCACCCCTGTTATTT 360
Qy 361 ATACAGATGTTGGGTATTTCTTCATCTTTTATATCTTGTACTTGAATTTTGTCTCTCTCA 420
Db 361 ATACAGATGTTGGGTATTTCTTCATCTTTTATATCTTGTACTTGAATTTTGTCTCTCTCA 420
Qy 421 TCGATGTTAAACAGGATCTATCTGTTATATATAGTTTCTTTGTCACATGATAGTATTTA 480
Db 421 TCGATGTTAAACAGGATCTATCTGTTATATATAGTTTCTTTGTCACATGATAGTATTTA 480
Qy 481 TTGATTCGTTACCCCTGCTCACTATTTTGTGATATACAGACCGAGGAAATTTAA 540
Db 481 TTGATTCGTTACCCCTGCTCACTATTTTGTGATATACAGACCGAGGAAATTTAA 540
Qy 541 CGAATTAATCAATAAAGACAGACCTAAAGTAAATGTTTAA 582
Db 541 CGAATTAATCAATAAAGACAGACCTAAAGTAAATGTTTAA 582

RESULT 2

US-09-134-001C-2402
; Sequence 2402, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 2402
; LENGTH: 615
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-2402

Query Match 18.6%; Score 108.2; DB 4; Length 615;
Best Local Similarity 50.7%; Pred. No. 6.7e-24;
Matches 309; Conservative 0; Mismatches 273; Indels 27; Gaps 1;

Qy 1 ATGTTAATGCTTTATTCATTTTTCGCTTACTTGTATAGGACGACATTCATCTGCGTTA 60
Db 7 ATGATGATCATCGTCATCTTAATCTTGATGTTATCTGATTCGTCATTCGCAAGCGGTTA 66
Qy 61 ATTGTGGCAAGCTTGCCAAAGGAATGATATTCGGGAGCACGGAAGCGCACTTAGGC 120
Db 67 ATATTGTTAAATTTATTTTAAAGATATATAGACATACGTTAGTGAATACTGGA 126
Qy 121 GCTACCAATGCATTCGTCATATGGGTGTAAAGCTGTTTGGTGTGTCATAGCGGAGAT 180
Db 127 GCAACTAACAGTTCGCTGTTCTTGGGAAGACGACGCTGATTTATAGTTACGTTTATAG 186
Qy 181 ATTTTGAAGGGACACTGGCAACTGCATTCGCTTTCTCATGCATGTTGATATTCACCG 240
Db 187 ATTTTGAAGGGATTTATACAGTCTTTTTCACATGTTGTTCCAGATTCATCGGATGGT 246
Qy 241 CTTCTT-----CGAGGAGTCTTTTGGGTTTATAGGCAC 273
Db 247 GTTATAGCACCTTCTTTACAAATGTTTAAATAGATGTTTGAATACTCGGTCAC 306
Qy 274 GTGTTTCCCATCTTCGCCAAATTTAAAGCGGTAAAGCCGTGGGACATCAGGAGCGGTT 333
Db 307 GTGATCCAAATATATCGAATTTTAAATGCGGGAAGACAGTACCTACCGAGGAGTT 366
Qy 334 TTGCTATTTTACGACCCCTGTTATTTATACAGATGTTGCGGTATTTCTTCATCTTTTAA 393
Db 367 GTATTAGGTGTCAAATCCTATTTTACTTCTTATCTTGGCAATATCTTTTATGTTATTA 426

Qy 394 TACTTGACTAAATTTCTTCTCTCATCGATGTTTAAAGGGATCTATCTACTGTTATATAT 453
Db 427 AATATCTTTAAATATGTTTCTTATCAAGTATCATTTGCAGCAATTTAGTGTGTCATGGT 486
Qy 454 AGTTCTTTTCCATGATACGATATTTTATTTGATTTGCTTACCTGCTCCTACTATTTTGTG 513
Db 487 TCAATCATCATTTTCATGATTTATTTTACTTGTCTGTTAGCGGAATTTGTTTCAATCATATTA 546
Qy 514 ATATACAGACCGGAGCAATTAACGAATTTATCAATAAAGACAGACCTTAAAGTAA 573
Db 547 ATAATTCGACACAAATCTAATATAGTTAGAATTTTAAAGGAGAGAAAGCACTTAAATTA 606
Qy 574 TGGTTATAA 582
Db 607 TGCATGTAA 615

RESULT 3

US-08-961-527-37
; Sequence 37, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6171 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-527-37

Query Match 17.2%; Score 100; DB 4; Length 6171;
Best Local Similarity 52.9%; Pred. No. 6.9e-21;
Matches 239; Conservative 0; Mismatches 210; Indels 3; Gaps 1;

Qy 1 ATGTTAATGCTTTATTTGCTTACTTGTATAGGACGACATTCATCTGCGTTA 60
Db 5149 ATGATTACAAATAGTATTTTATTAATCTAGCCTATCTGCTGGTTCGATTCATCTGCTC 5208
Qy 61 ATTCTGGCAAGCTTCCCAAGCAATTTGATTTCCGGAGCACGGAAGCGCACTTAGGC 120
Db 5209 TGAATTTGGCAAGATTTCTTTCAATCAATCTACGCGAGCATGTTCTTGGTAACACTGGA 5268
Qy 121 GCTACCAATCATTCCTGCTACATTTGGGTGTAAAGCTGTTTCGCTCGTCATAGCGGAGAT 180

```

Db      306   3TTTCACTGTCAGTGTCACAGCATCGATCCGGCTGTGA 344

RESULT 5
US-08-232-463-14
; Sequence 14: Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOMLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:

```

```

, OPERATING SYSTEM: PC-DOS/MS-DOS
, SOFTWARE: PatentIn Release #1.0, Version #1.25
, CURRENT APPLICATION DATA:
, APPLICATION NUMBER: US/08/935,313
, FILING DATE:
, CLASSIFICATION: 435
, PRIOR APPLICATION DATA:
, APPLICATION NUMBER: US/07/935,313
, FILING DATE:
, APPLICATION NUMBER: EP 91 114 300.6
, FILING DATE: 26-AUG-1991
, ATTORNEY/AGENT INFORMATION:
, NAME: BENT, Stephen A.
, REGISTRATION NUMBER: 29,768
, REFERENCE/DOCKET NUMBER: 30472/114 IMMU
, TELECOMMUNICATION INFORMATION:
, TELEPHONE: (703)836-9300
, TELEFAX: (703)683-4109
, TELEX: 899149
, INFORMATION FOR SEQ ID NO: 14:
, SEQUENCE CHARACTERISTICS:
, LENGTH: 7218 base pairs
, TYPE: nucleic acid
, STRANDEDNESS: single
, TOPOLOGY: linear
, IMMEDIATE SOURCE:
, CLONE: ptz9pt-f15
, PS-08-232-463-14

```

Qy	113	ACATAGGCGCTACCAATATGCATTTGGTCTATATGGGTGTAAAGCTGGTTCGGTCTGCTATATG	17
Db	6	ACACTGGAACGACCAACACCTCCGCATTTTAGTGAAGAACTGTAATGGCAACCTTTG	65
Qy	173	CCGGAGATATTTTGAAGGACACTGGCAACTGCATTCCTTTCTCATCGCATGTTGATA	232
Db	66	TGATTTGACTTTTTTCAAGGAACCTAGCAACGCTGCTCCGATTAATTTTTCATCTACAAG	125
Qy	233	---TTCACCCGCTCTTGCAGGAGCTTTCCGGTTTTAGGCCACGTGTTTCCCATCTTCG	289
Db	126	CGGTTCTCTCTCATCTTTGGACATTTTGGCTGTATFCGGCCATACCTTCCCTCATCTTTG	185
Qy	290	CCAAATTTAAAGCGGTAAAGCGCTGGCGACATCAGSAGGCGTTTTCGTATTTTACGAC	349
Db	186	CAGGATTTAAAGGTGTAAAGCTGTCGCAACACAGTGTCTGGAGTGATTTTCGGATTTGCGC	245
Qy	350	CCCTGTTATTTATCAGCATGTTCCGGTATTCCTCATCTTTTATACTTCGACTAAATTTG	409
Db	246	CTATCTCTGCTCTACTTTGCGATTAATCTCTTTGGACTCTCATATCTTGGCAGTATGA	305
Qy	410	TTTCTCTCTCATCGATGTTAACAGGGATCTATACGTGA	448

Query Match	8.4%	Score 48.8	DB 1	Length 7218		
Best Local Similarity	2.5%	Pred. NO. 4.2e-05				
Matches	8	Conservative 191	Mismatches 123	Indels 0	Gaps 0	
Qy 189	AGGGACACTGCCA	CTGCATTGCC	TTTTCTCAT	GCATCTTG	TATATTCACCGCTCTTGC	248
Db 1053	AGGGACCTTCG	ATGATATG	ATGATATG	ATGATATG	ATGATATG	1112
Qy 249	AGGAGCTTTCGG	TTTTAGCC	ACGGTGTTC	CCCATCTT	CGCCAAATTTA	308
Db 1113	YYYYYYYYYYYY	YYYYYYYY	YYYYYYYY	YYYYYYYY	YYYYYYYY	1172
Qy 309	AGCGTGCGCAC	ATCAGGAG	CGGTTTGCT	ATTATTAC	GACCCCTGTTATTAT	368
Db 1173	YYYYYYYYYYYY	YYYYYYYY	YYYYYYYY	YYYYYYYY	YYYYYYYY	1232
Qy 369	GGTTCGGGTAT	CTTCATCT	TTTTATAC	TCTGAC	TAAATTTGTTCTCTCAT	428
Db 1233	YYYYYYYYYYYY	YYYYYYYY	YYYYYYYY	YYYYYYYY	YYYYYYYY	1292
Qy 429	AACAGGGATC	TACTGTAT	ATATAG	TTCTTTCT	GCATCATACG	488

```

Db 1293 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1352
QY 489 CGTACCCTGCTCACTATTTT 510
Db 1353 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1374

RESULT 6
US-08-487-826B-13/c
; Sequence 13, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19124 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-487-826B-13

Query Match 6.4%; Score 37; DB 2; Length 19124;
Best Local Similarity 57.3%; Pred. No. 0.29;
Matches 67; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 377 TATCTTCATCTTTTATACCTGACGATAATTTGTTCTCTCATCGATGTTAACAGGA 436
Db 15907 TTTTAAATTTTATTTTAAATTTTATTTTAAATTTTATTTTATTTTATTTTAAAT 15848

QY 437 TCTATACCTGTATATAGCTTTCTTGTCACGATACGATATTATTTGATTCGTTA 493
Db 15847 TAAATTTTATTTATTTATTTTATTTTATTTTAAATAAATTTTATTTTATTTATGTA 15791

RESULT 7
US-09-181-336-14
; Sequence 14, Application US/09181336A
; Patent No. 6242238

```

FILING DATE: 18-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/672,183
FILING DATE: 20-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Frommer, William S.
REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454310-2570
TELEPHONE: (212) 840-3333
TELEFAX: (212) 840-0712
TELEX: 425066 CURTMS
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 1320 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-257-073-15

Query Match: 5.9%; Score 34.2; DB 1; Length 1320;
Best Local Similarity 48.7%; Pred. No. 0.58;
Matches 93; Conservative 0; Mismatches 98; Indels 0; Gaps 0;
Qy 332 TTTTCTATTATTTACGACCCCTGTTATTTATCATCAGATGGTTGCGGTATTTCTCATCTTTT 391
Db 812 TTTTCTATGTCAGCAGATTATCTCTTTTCTGTTGTTTATTGAACCATCTTTCTTT 753
Qy 392 TATACCTGACTAAATTTGTTCTCTCATCATGATGTTAACAGGGATCTATCTGTTATAT 451
Db 752 TCTTCAAGATGTCCTTTATGTATATATCCCTTCCTTCAACATTTGTTGTAATAGATTCT 693
Qy 452 ATAGTTTCTTCTCCATGATACGTATTTATTCATGTTCTGTTACCCCTGCTCACTATTTTG 511
Db 692 CTATTTGTTTCTTATATAGATATTTCCCTTGGATCTCTACTATTTCCCTCTGTTGCC 633
Qy 512 TCATATACAGA 522
Db 632 TCAGGTAAGA 622

RESULT 9
US-08-098-327E-41/c
Sequence 41, Application US/08098327E
Patent No. 6270771
GENERAL INFORMATION:
APPLICANT: GUERIN-MARCHAND, Claudine
APPLICANT: DRUILHE, Pierre
TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
TITLE OF INVENTION: HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
TITLE OF INVENTION: OF STIMULATING THE T LYMPHOCYTES
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/098,327E
FILING DATE: 24-NOV-1993
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 91 01286
FILING DATE: 05-FEB-1991
ATTORNEY/AGENT INFORMATION:
NAME: McGowan, Malcolm K.

REGISTRATION NUMBER: 39,300
REFERENCE/DOCKET NUMBER: 010830-045
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 41:
SEQUENCE CHARACTERISTICS:
LENGTH: 1482 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1482
PUBLICATION INFORMATION:
DOCUMENT NUMBER: WO 92/13884
PUBLICATION DATE: 20-AUG-1992
US-08-098-327E-41

Query Match: 5.9%; Score 34.2; DB 4; Length 1482;
Best Local Similarity 48.7%; Pred. No. 0.61;
Matches 93; Conservative 0; Mismatches 98; Indels 0; Gaps 0;
Qy 332 TTTTGTATTATTTACGACCCCTGTTATTTATCATCAGATGGTTGCGGTATTTCTCATCTTTT 391
Db 974 TTTTGTATGTCAGCAGATTATCTCTTTTCTGTTGTTTATTGAACCATCTTTCTTT 915
Qy 392 TATACCTGACTAAATTTGTTCTCTCATCATGATGTTAACAGGGATCTATCTGTTATAT 451
Db 914 TCTTCAAGATGTCCTTTATGTATATATCCCTTCCTTCAACATTTGTTGTAATAGATTCT 855
Qy 452 ATAGTTTCTTCTCCATGATACGTATTTATTTATGTTGTTACCCCTGCTCACTATTTTG 511
Db 854 CTATTTGTTTCTTATATAGATATTTCCCTTGGATCTCTACTATTTCCCTCTGTTGCC 795
Qy 512 TCATATACAGA 522
Db 794 TCAGGTAAGA 784

RESULT 10
US-08-462-625-41/c
Sequence 41, Application US/08462625
Patent No. 6319502
GENERAL INFORMATION:
APPLICANT: GUERIN-MARCHAND, Claudine
APPLICANT: DRUILHE, Pierre
TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
TITLE OF INVENTION: HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
TITLE OF INVENTION: OF STIMULATING THE T LYMPHOCYTES
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Burns, Doane, Swecker & Mathis
STREET: P.O. Box 1404
CITY: Alexandria
STATE: Virginia
COUNTRY: United States
ZIP: 22313-1404
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/462,625
FILING DATE: 05-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/098,327
FILING DATE: 24-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 91 01286

```

Query Match      5.9%; Score 34.2; DB 4; Length 1493;
Best Local Similarity 48.7%; Pred. NO. 0.60;
Matches 93; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

```

	Matches	93; Conservative	0; Mismatches	98; Indels	0; Gaps
Qy	332	TTTTTGCTATTTTACGACACCCCTGTTATTTATCATCAGATGGTTGGGTATTCCTTCATCTTTT	391		
Db	971	TTTTGTGATGTCAGCAGATTATCTCTTTTGTGCTGGTTTTATTGAACCATCTCTTTCTTT	912		
Qy	392	TATACTTGACTAAATTTGTTTCTCTCATCGATGTTTAAACAGGATCTATACGTGTTATAT	451		
Db	911	TCTTCAAGATGTCCTTTATGTATATCCCTTCGCTCTCAACATTTGTTGTAATAGATTC	852		
Qy	452	ATAGTTCTTTGTGCCATGATACGATATTTATTTGATTCGTGTACCCCTGCCTCAGCTATTTTG	511		
Db	851	CTATTTGTTTTTCTATATATAGATATTTCCCTCGAATCTCTACTATTCCTCTGTTGTCC	792		
Qy	512	TGATATACAGA	522		
Db	791	TGAGGTAAGA	781		

RESULT 12
US-08-462-625-38/c
: Sequence 38, Application US/08462625
: Patent No. 6319502
: GENERAL INFORMATION:
: APPLICANT: GUERIN-MARCHAND, Claudine
: APPLICANT: DRUILHE, Pierre
: TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
: TITLE OF INVENTION: HEPATIC STAGES OF P. FALCIPARUM BEARING EPITOPES CAPABLE
: TITLE OF INVENTION: OF STIMULATING THE T LYMPHOCYTES
: NUMBER OF SEQUENCES: 46
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Burns, Doane, Swecker & Mathis
: STREET: P.O. Box 1404
: CITY: Alexandria
: STATE: Virginia
: COUNTRY: United States
: ZIP: 22313-1404
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/462,625
: FILING DATE: 05-JUN-1995
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/098,327
: FILING DATE: 24-NOV-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: FR 91 01286

Query Match 5.8%; Score 34; DB 4; Length 950;
Best Local Similarity 50.0%; Pred. No. 0.57;
Matches 85: Conservative 0; Mismatches 85: Indels

RESULT 14
US-08-462-625-32/c
; Sequence 32, Application US/08462625
; Patent No. 6319502
; GENERAL INFORMATION:

TITLE OF INVENTION: PEPTIDE SEQUENCES SPECIFIC FOR THE
 TITLE OF INVENTION: HEPATIC STAGES OF P. FALCIPARUM BEARING EPTIOPES CAPABLE
 TITLE OF INVENTION: OF STIMULATING THE T LYMPHOCYTES
 NUMBER OF SEQUENCES: 46
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Burns, Doane, Swecker & Mathis
 STREET: P.O. Box 1404
 CITY: Alexandria
 STATE: Virginia
 COUNTRY: United States
 ZIP: 22313-1404
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/462,625
 FILING DATE: 05-JUN-1995
 CLASSIFICATION: 424
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/098,327
 FILING DATE: 24-NOV-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: FR 91 01286
 FILING DATE: 05-FEB-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: McGowan, Malcolm K.
 REGISTRATION NUMBER: 39,300
 REFERENCE/DOCKET NUMBER: 010830-078
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 836-6620

TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 950 base pairs
TYPE: nucleic acid
STRADEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
PUBLICATION INFORMATION:
DOCUMENT NUMBER: WO 92/13884
PUBLICATION DATE: 20-AUG-1992
US-08-462-625-32

Query Match 5.8%; Score 34; DB 4; Length 950;
Best Local Similarity 50.0%; Pred. No. 0.57; Indels 0; Gaps 0;
Matches 85; Conservative 0; Mismatches 85; Indels 0; Gaps 0;
QY 353 TGTATTATATACGATGTTGGGATTTCTTCATCTTTTATATACCTGACTAAATTTGTTT 412
DB 943 TCTTCTTTTGTCTCGTTTATTGAACCATCTTCTTCTCAAGATGCTCTTATGT 884
QY 413 CTCTCTCATCGATGTTAAACAGGATCTATATCTGTATATATAGTTTCTTTGTCATGATA 472
DB 883 ATATCCCTTCGCTCTCAACATTTGTTGTAATAGATTCCTATTTGTTTTCTATATA 824
QY 473 CGTATTATTATGTTGCTTACCCTGCTCCTATTTTCTGATATACAGA 522
DB 823 GATATTTCCTGGAATCTCTACTATTTCCCTCTGTTGCTGAGGTAAGA 774

RESULT 15
US-09-128-155-16/C
Sequence 16, Application US/09128155
Patent No. 6117634
GENERAL INFORMATION:
APPLICANT: Pan, Yang
TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
FILE REFERENCE: 09404/052001
CURRENT APPLICATION NUMBER: US/09/128,155
CURRENT FILING DATE: 1998-08-03
EARLIER APPLICATION NUMBER: US 60/091,650
EARLIER FILING DATE: 1998-07-02
EARLIER APPLICATION NUMBER: US 60/054,646
EARLIER FILING DATE: 1997-08-04
NUMBER OF SEQ ID NOS: 18
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 16
LENGTH: 152331
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(152331)
OTHER INFORMATION: n = A,T,C or G
US-09-128-155-16

Query Match 5.8%; Score 33.8; DB 3; Length 152331;
Best Local Similarity 49.7%; Pred. No. 7.7;
Matches 86; Conservative 0; Mismatches 87; Indels 0; Gaps 0;
QY 390 TTTATACCTTGACTAAATTTGTTCTCTCATCGATGTTAAACAGGATCTATACGTAT 449
DB 37151 TTCTTCTTACCTAAATTTATTCTTTAAGCTTCTTACCAAAAAATACCACCTTATT 37092
QY 450 ATATAGTTTCTTTGTCATGATACGATTTATTGATTTGCTTACCCTGCTCCTCATTTT 509
DB 37091 TCTATAATTTCTTTACATTTCCCTTATTCTCTGTTCTCTTACCTTTGTTTATACATA 37032
QY 510 TGTGATATACAGACCGAGCGAACAATTAACGAATTAATCAATAAAACAGAAC 562
DB 37031 ACTTTTCAATAAACTTGAATTAGACAAAAATTTATTCACCTTTTAAAAAGGAC 36979

Search completed: December 25, 2002, 14:26:19
Job time : 135.875 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 25, 2002, 11:18:11 ; Search time 171.89 Seconds
(without alignments)
1375.466 Million cell updates/sec

Title: US-10-068-080-4
Perfect score: 582
Sequence: 1 atgttaattgtttattgat.....ctaaagtaaatgtttataa 582

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 363474 seqs, 203117208 residues
Total number of hits satisfying chosen parameters: 726948

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	582	100.0	582	12	US-10-068-080-4
c	2	109.6	18.8	6691	Sequence 4, Appli
3	101.6	17.5	642	10	US-09-070-927A-88
4	101.6	17.5	642	10	US-09-815-242-9231
5	100.7	17.2	642	12	US-09-815-242-9438
6	90.6	15.6	609	10	US-10-068-080-2
7	90.6	15.6	609	10	US-09-823-246-1
8	88.2	15.2	606	10	US-09-815-242-8039
9	48	8.2	600	10	US-09-815-242-4250
10	46.4	8.0	72	10	US-09-815-242-6900
11	43	7.4	789	10	US-09-974-300-4100
12	42.4	7.3	612	10	US-09-815-242-7356
13	40.2	6.9	449	10	US-09-815-242-9659
14	39.8	6.8	663	10	US-09-960-352-2412
15	38	6.5	397658	10	US-09-815-242-7517
16	37.8	6.5	618	10	US-09-813-320-3
c	17	37.2	6.4	6786	Sequence 3, Appli
18	36.6	6.3	300	10	US-09-815-242-6206
c	19	36.6	6.3	312	Sequence 72, Appli
					Sequence 6251, Ap
					Sequence 8414, Ap

ALIGNMENTS

RESULT 1
US-10-068-080-4
: Sequence 4, Application US/10068080
: Patent No. US20020115591A1
: GENERAL INFORMATION:
: APPLICANT: Youngman, Philip
: TITLE OF INVENTION: USE OF YNES, ESSENTIAL BACTERIAL GENES AND POLYPEPTIDES
: FILE REFERENCE: 07334/090001
: CURRENT APPLICATION NUMBER: US/10/068,080
: CURRENT FILING DATE: 2002-02-05
: PRIOR APPLICATION NUMBER: US 60/070,116
: PRIOR FILING DATE: 1997-12-31
: NUMBER OF SEQ ID NOS: 12
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 4
: LENGTH: 532
: TYPE: DNA
: ORGANISM: Streptococcus pneumoniae
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (1)...(579)
US-10-068-080-4

Query Match	100.0%	Score 582;	DB 12;	Length 582;
Best Local Similarity	100.0%	Pred. No. 1.7e-152;		
Matches 582;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Qy 1 ATGTTAATTCGCTTATTGATTTTGGCCCTACTTGATAGCAGCATTCCTATCGGCTTA 60
Db 1 ATGTTAATTCGCTTATTGATTTTGGCCCTACTTGATAGCAGCATTCCTATCGGCTTA 60
Qy 61 ATTCGGCAGGCTGCCAAGGAATTCATATTCGGGAGCAGCAAGCGCACTTAGGC 120
Db 61 ATTCGGCAGGCTGCCAAGGAATTCATATTCGGGAGCAGCAAGCGCACTTAGGC 120
Qy 121 GCTACCAATGATTCCTGCTACATTTGGTGTAAAGCTGGTTCGGTCGTCATAGCCGGAGAT 180
Db 121 GCTACCAATGATTCCTGCTACATTTGGTGTAAAGCTGGTTCGGTCGTCATAGCCGGAGAT 180
Qy 181 ATTTTGAAGGACATGGCAACTGCATTTGCTTTCTCATSCATGTTGATATTACCCG 240
Db 181 ATTTTGAAGGACATGGCAACTGCATTTGCTTTCTCATSCATGTTGATATTACCCG 240

20	36.6	6.3	70768	9	US-10-135-322-13	Sequence 13, Appl	
c	21	35.8	6.2	425	10	US-09-834-975-451	
c	22	35.6	6.1	424	10	US-09-960-352-11218	
c	23	35.4	6.1	375	10	US-09-960-352-15014	
24	35.2	6.0	453	10	US-09-960-352-7418	Sequence 7418, Ap	
25	35.2	6.0	774	10	US-09-910-943-506	Sequence 506, Appl	
26	35.2	6.0	99916	10	US-09-816-095-3	Sequence 3, Appl	
27	34.8	6.0	397	10	US-09-960-352-13784	Sequence 13784, A	
28	34.8	6.0	523	10	US-09-864-761-30606	Sequence 30606, A	
29	34.6	5.9	414	10	US-09-960-352-6528	Sequence 6528, Ap	
30	34.6	5.9	1503841	9	US-09-946-807-1	Sequence 1, Appl	
31	34.6	5.9	1503841	10	US-09-795-668-1	Sequence 1, Appl	
32	34.6	5.9	1503841	10	US-09-795-668-1	Sequence 2223, Ap	
c	33	34.4	5.9	415	10	US-09-960-352-2223	
c	34	34.4	5.9	1198	10	US-09-789-561-55	Sequence 55, Appl
35	34.4	5.9	66686	10	US-09-736-960-86	Sequence 86, Appl	
36	34.2	5.9	598	10	US-09-864-761-14959	Sequence 14959, A	
37	34.2	5.9	598	10	US-09-864-761-31457	Sequence 31457, A	
38	34.2	5.9	995	10	US-09-070-927A-833	Sequence 833, Appl	
c	39	34.2	5.9	1482	10	US-09-837-344-41	
c	40	34.2	5.9	1493	10	US-09-837-344-38	Sequence 38, Appl
41	34.2	5.9	1580	10	US-09-070-927A-347	Sequence 347, Appl	
42	34.2	5.9	1920	10	US-09-800-065-1	Sequence 1, Appl	
43	34.2	5.9	5513	10	US-09-800-065-3	Sequence 3, Appl	
c	44	34.2	5.9	21724	10	US-09-764-864-1603	
45	34.2	5.9	21727	10	US-09-764-864-1604	Sequence 1604, Ap	

Db 181 ATTTGAAAGGACACTGGCAACTGCATGCTTTCTCATGCATGTTGATATTCACCCG 240
Qy 241 CTTCTTCGACAGCTCTTTGGCGTTTATAGGCACACGTTTCCCATCTTCGCCAAATTTAAA 300
Db 241 CTTCTTCGACAGCTCTTTGGCGTTTATAGGCACACGTTTCCCATCTTCGCCAAATTTAAA 300
Qy 301 GCGGTAAAGCGTGGCGACATCAGGAGCGTTTGTGCTATTTTACGACACCCCTCTTATTT 360
Db 301 GCGGTAAAGCGTGGCGACATCAGGAGCGTTTGTGCTATTTTACGACACCCCTCTTATTT 360
Qy 361 ATCAGATGTTGGGGTATCTTCATCTTTTATATATATATAGTTCTTTGTCATGATACATTTA 480
Db 361 ATCAGATGTTGGGGTATCTTCATCTTTTATATATATATAGTTCTTTGTCATGATACATTTA 480
Qy 421 TCGATGTTAAAGGATCTATCTGTTATATATATATAGTTCTTTGTCATGATACATTTA 480
Db 421 TCGATGTTAAAGGATCTATCTGTTATATATATATAGTTCTTTGTCATGATACATTTA 480
Qy 481 TTGATGTCGTTACCTCTCTCACTATTTTGTGATATACACACCGAGCGAACTATAA 540
Db 481 TTGATGTCGTTACCTCTCTCACTATTTTGTGATATACACACCGAGCGAACTATAA 540
Qy 541 CGAATTATCAATAAAGACACCTAAAGTAAATGGTTATAA 582
Db 541 CGAATTATCAATAAAGACACCTAAAGTAAATGGTTATAA 582

RESULT 2
US-09-070-927A-88/c
; Sequence 88, Application US/09070927A
; Patent No. US20020120116A1
; GENERAL INFORMATION:
; APPLICANT: Charles A. Kunsch
; Steven Barash
; Patrick J. Dillon
; TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 982
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/070,927A
; FILING DATE: 04-May-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/046,655
; FILING DATE: 1997-05-16
; APPLICATION NUMBER: 60/044,031
; FILING DATE: 1997-05-06
; APPLICATION NUMBER: 60/066,009
; FILING DATE: 1997-11-14
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB369
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 88:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6691 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 88:
US-09-070-927A-88
Query Match 18.8%; Score 109.6; DB 10; Length 6691;
Best Local Similarity 53.4%; Pred. No. 8.9e-21;
Matches 254; Conservative 0; Mismatches 219; Indels 3; Gaps 1;
Qy 1 ATGTTAATCTCTTATTTGATTATTTTGGCTTACTTTGATAGACGACATCTCCATCTGGCTTA 60
Db 5424 ATGAAAATCGTCATTTTGTACTTGTTCCTTATTTATAGTTGCTATCTCCCTCAGGTGTT 5365
Qy 61 ATTGTGGCAAGCTTGCACAAAGGAATTTGATATTCGGGACGACGGAAGCGCAACTTAGGC 120
Db 5364 TGGATTGGTAAACTTTTCTTTTAAAGAGATATACGCCAATTTGGGAGTGGGAATACAGGA 5305
Qy 121 GCTACCAATGATCTCCATACATTTGGGTGTTAAAGCTGGTTCGGTCGTCATAGCCGGAGAT 180
Db 5304 ACACCAATACATTTGCTGCTTTAGGGAACCTCGCGAATTTACGGTATTTAATGAT 5245
Qy 181 ATTTTGAAGGACACTGGCAACTGCATTCCTTTTCTCATGCATGT--TGATATTAC 237
Db 5244 ATCTTGAAGGAACGTTAGCCACTTCATACCTTATTTGTTGTTTACAGGCGTGAAT 5185
Qy 238 CCGCTTCTTCGACGAGTCTTTGCGGTTTATAGCCACGTTTCCCATCTTCGCCAAATTT 297
Db 5184 CCGCTCTCTTTGGGTAGCAGCTGTTTATAGGCGATACCTTCCCTATTTTGGCAATTC 5125
Qy 298 AAGGCGGTAAAGCGTGGGACATCAGGAGCGCTTTTGTCTATTTTATAGCACCCCTGTTA 357
Db 5124 AAGGCGGTAAAGCGTGGGACATCAGGAGCGCTGCTGCTATTTATAGCATACAGCCCAATTT 5065
Qy 358 TTTATACAGATGTTGCGGTATTTCTTCATCTTTTATACCTTATTTTACCTAAATTTGTTCTC 417
Db 5064 TTTATTTATCTGCTCTTATTTTGTGATTTGCTGCTATCTATCTAACCAATGTTGAGTTA 5005
Qy 418 TCATCATGTTTAAAGGAGTCTATATCTGTTATATATATAGTTTCTTGTCCATGATAC 473
Db 5004 ACAAGTATGATTAGCGCTGCTACTAATTAATACACTTTCTACTATATTTTGGCTTTTTCAC 4949

RESULT 3
US-09-815-242-9231
; Sequence 9231, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110

```

Db 361 TGTCTCTACCTTGGAGTATCTCTTTGGAGCTCTCTATCTTGGCAGTATGATTTCACGTG 420
Qy 418 TCATCGATGTTAAACAGGAGTCATATACCTGTTAT 449
    | | | | | | | | | | | | | | | | | | | |
Db 421 TCTAGTGTCAACAGCATCGATCGCGGCTGTTAT 452
    | | | | | | | | | | | | | | | | | | | |
RESULT 5
US-10-068-080-2
; Sequence 2, Application US/10068080
; Patent No. US20020115591A1
; GENERAL INFORMATION:
; APPLICANT: Fritz, Christian
; APPLICANT: Youngman, Philip
; TITLE OF INVENTION: USE OF YNES, ESSENTIAL BACTERIAL GENES AND POLYPEPTIDES
; FILE REFERENCE: 07334/090001
; CURRENT APPLICATION NUMBER: US/10/068,080
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/070,116
; PRIOR FILING DATE: 1997-12-31
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 642
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae

```

FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(639)
US-10-068-080-2

Query Match 17.2%; Score 100; DB 12; Length 642;
Best Local Similarity 52.9%; Pred. No. 1.6e-18;
Matches 239; Conservative 0; Mismatches 210; Indels 3; Gaps 1;

QY 1 ATGTTAATTCCTTATGATTAATTTGGCTACTTCCATGATAGGACGAGCATTCATCGCTTA 60
DB 1 ATGTTAATTCCTTATGATTAATTTGGCTACTTCCATGATAGGACGAGCATTCATCGCTTA 60
QY 61 ATTGTGGCAAGCTTGCCAAAGAAATGATATTCCGGGAGCAGCAAGCGGCAACTTAGGC 120
DB 61 TGGATTGGCAAGTATTCCTTCAATCAATCTACGGGAGCATGTTCTGGTAACACTGGA 120
QY 121 GCTACCAATGCATTCGGTACATTTGGGTGTAAGAGCTGGTTCGTCGTCATAGCCGGAGAT 180
DB 121 ACACCAACACCTTCGCAATTTTAGGTAAAGAAAGCTGGTATGCAACCTTTGTGATTGAC 180
QY 181 ATTTTGAAGGGACACCTGGCAACTGCATTCGCCCTTTTCTCATGATGT---TGATATTAC 237
DB 181 TTTTTCAGAGAACCCCTAGCAAGCTGCTTCCGATTATTTTTCATCTACAAAGCGGTTCT 240
QY 238 CCGCTTCTTCAGAGGCTTTGGGTTTAAAGCCACGCTTTCCCATCTTCGCCAAATTT 297
DB 241 CCTCTCATCTTTGGCTGTTATCGGCCATACCTTCCCTATCTTTGCGAGGATTT 300
QY 298 AAGGCGGTAAAGCGTGGCGACATCAGGAGGCGTTTGGCTATTTTACGCCACCCCTGTTA 357
DB 301 AAGGCGGTAAAGCGTGGCGACATCAGGAGGCGTTTGGCTATTTTACGCCACCCCTGTTA 360
QY 358 TTTATACAGAGGTGGCGTATCTTCATCTTTTATACGTTGACTAAATTTGTTCTCTC 417
DB 361 TGTCTTACCTTGGATTATCTTCTTTGGAGCTCTCTATCTTGGCAGTATGATTTCACTG 420
QY 418 TCATCGATGTTAAAGGATCTATACGTTAT 449
DB 421 TCTAGTCTACAGATCGATTGGCGCTGTTAT 452

RESULT 6

US-09-823-246-1
Sequence 1, Application US/09823246
Patent No. US2002058789A1
GENERAL INFORMATION:
APPLICANT: Burnham, Martin K. R.
APPLICANT: Biswas, Sanjoy
APPLICANT: Zalacain, Magdalena
APPLICANT: Warren, Patrick V.
APPLICANT: Slyvester, Damien
TITLE OF INVENTION: ynes
FILE REFERENCE: GM20001
CURRENT APPLICATION NUMBER: US/09/823,246
PRIOR FILING DATE: 2001-03-30
PRIOR APPLICATION NUMBER: US 60/193,496
NUMBER OF SEQ ID NOS: 2
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 609
TYPE: DNA
ORGANISM: Staphylococcus aureus
US-09-823-246-1

Query Match 15.6%; Score 90.6; DB 10; Length 609;
Best Local Similarity 48.9%; Pred. No. 6.3e-16;
Matches 298; Conservative 0; Mismatches 284; Indels 27; Gaps 1;

QY 1 ATGTTAATTCCTTATGATTAATTTGGCTACTTCCATGATAGGACGAGCATTCATCGCTTA 60
DB 1 ATGTTAATTCCTTATGATTAATTTGGCTACTTCCATGATAGGACGAGCATTCATCGCTTA 60

DB 1 ATGATGATAATCGTCATGTTACTACTAAGTATTTATCTTATCGGCGCTTTCCCAAGTGGATTC 60
QY 61 ATTGTGGCAAGCTTGCCAAAGAAATGATATTCCGGGAGCAGCAAGCGGCAACTTAGGC 120
DB 61 GTPAATTTGGAAAAATTAATTTTCAAAAAAGATATTAGACAAATTTGGTAGTGGTAATACTGGC 120
QY 121 GCTACCAATGCATTCGGTACATTTGGGTGTAAGAGCTGGTTCGTCGTCATAGCCGGAGAT 180
DB 121 GCTACTAATAGCTTTAGAGTATTTAGTCTGCTCGCAGGATTTCTTGTAAACATTTCTTAGAT 180
QY 181 ATTTTGAAGGGACACCTGGCAACTGCATTCGCCCTTTTCTCATGATGTGA----- 230
DB 181 ATTTTCAAGGGTTCATAACTGTTTCTTCCCTTTATGGTTACAACTTCACGCGAGATGGC 240
QY 231 -----TATTACCCGCTTCTTGCAGGAGTCTTTGCGGTTTTAGGCCAC 273
DB 241 CTTATTAGTACTTTTTTTTACAAATGTTTAAATTTGTTGGCTTATTCGTATATCTTGGACAC 300
QY 274 GTGTTTCCCATCTTCGCCAAATTTAAAGCGGTAAAGCGCTGGCGACATCAGGAGCGTT 333
DB 301 GTTATCTCTGTTTAAATTTCCAAAGTGGCAAGCTGTTGCAACTAGTGCAGGTGTC 360
QY 334 TTGCTATTTTACGACCCCTGTTTATTTATCAGATGTTTCCGCTATTTCTCATCTTTT 393
DB 361 GTCCTGGAGTCAATCCGATCTTTTACTAATATCTTCAATTTATCTTCTTATGTTATG 420
QY 394 TACTTCACTAAATTTGTTCTCTCTCATCGATGTTAACAGGATCTATCTGTTATATAT 453
DB 421 AAGATTTTAAATATGTTTCTTTTAGCAAGTATCGTTGCAAGCTGTTGCTGTTGTTG 480
QY 454 AGTTTCTTTGTCATGATAGTATTTATTTGATGTTGTTACCTGCTGCTACTATTTT 513
DB 481 TCGCTATCATTCAGACATATATTTTATTTAGTCTGTTAGTTCTTCAATCATATTTG 540
QY 514 ATATACAGACCGGAGCAACATTTAAACGAATTTATCAATAAAACAGCACTTAAAGTAAA 573
DB 541 AATAATAGACATCGCTCTAATATCGCAAGATTTTTAGAGGCGCAAGCACTTAAATAAA 600
QY 574 TGGTATAA 582
DB 601 TGGATCTAA 609

RESULT 7

US-09-815-242-8039
Sequence 8039, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in Prokaryotes
FILE REFERENCE: ELITRA.011a
CURRENT APPLICATION NUMBER: US/09/815,242
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22

```
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8039
; LENGTH: 609
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; NAME/KEY: CDS
; FEATURE:
; LOCATION: (1)...(609)
US-09-815-242-8039

Query Match      15.6%; Score 90.6; DB 10; Length 609;
Best Local Similarity 48.9%; Pred. No: 6.3e-16;
Matches 298; Conservative 0; Mismatches 284; Indels 27; Gaps 1;

Qy 1 ATGTTAATGCTTTATGATTAATTTGGCCCTACTGATAGGACGATTCATCTGGCCTTA 60
Db 1 ATGATGATAATCGTCATCTACTACTAAGTTATCTTATCGGCGCTTTCCCAAGTGGATTC 60
Qy 61 ATTTGGGCAAGCTTGGCCAAAGGATTCATATTCGGGAGCACGGAAGCGGCAACTTAGGC 120
Db 61 GTAAATGGAAATATATTTTCAAAAAGATATATAGACAATTTGGTAGTGGTAATCTAGTC 120
Qy 121 GCTACCAATGCAATCCGTACATTTGGGTGTAAGAGCTGCTCGGTGCATAGCCGGAGAT 180
Db 121 GCTACTAATAGCTTTAGAGTATAGGTGCTCTCGCAGGATTCCTGGTAACATTTCTAGAT 180
Qy 181 ATTTTGAAGGACACTGGCAACTGCATGCTTTTCTCATGCAATGTTGA----- 230
Db 181 ATTTTCAAAAGGTTTCATAACTGTTTCTTCCCTTTATGTTTACCAGTTTCACGCGAGATGC 240
Qy 231 -----TATTCACCCGCTTCTGACGAGTCTTTGCGGTTTATAGGCCAC 273
Db 241 CCTATTAGTACTTTTTTACAAATGGTTAATTTGGCTTATTCGCTTACTTGGACAC 300
Qy 274 GTGTTTCCCATCTTCGCCAAATTTAAAGCGGTAAAGCGGTGGGACATCAGGAGCGTT 333
Db 301 GTTATCTCTGTTTATTTAAATTTCAAGGTGGCAAGCGGTGCAACTAGTGCAGGTGTC 360
Qy 334 TTGCTATTTTACGACCCCTGTTATTTATCAGATGGTTGCGGTATTTCTCATCTTTTA 393
Db 361 GTCTTGGGAGTCAACCCGATCTTTTACTAATCTTGAATATCTTCTTTATTGTTATG 420
Qy 394 TACTTGACTAAATTTGTTTCTCTCATCGATGTTTAAACAGGATCTACTGTTATATAT 453
Db 421 AAGATTTTAAATATGTTTCTTTAGCAAGTATCGTTGACGCAANTTGGTGTGATGGC 480
Qy 454 AGTTTCTTTGTCATGATACGTAATTTATTTGTTTACCCCTGCTCACCATTATTTTGT 513
Db 481 TCGCTTATCATCAAGACTATATTTTATTAGTCGTTAGTTTCTTAGTTCAATCATATTG 540
Qy 514 ATATACACACACCGGACGCAATTTAAACGAATTTATCAATAAACAAGACCTTAAAGTAAA 573
Db 541 ATAATTAGACATCGCTCTAATATCTCAAGAATTTTATAGAGCGGAAGAACCTTAAATAA 600
```

RESULT 8

```
US-09-815-242-4250
; Sequence 4250, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
```

```
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4250
; LENGTH: 606
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-815-242-4250
```

```
Query Match      15.2%; Score 88.2; DB 10; Length 606;
Best Local Similarity 48.8%; Pred. No. 2.9e-15;
Matches 295; Conservative 0; Mismatches 283; Indels 27; Gaps 1;
```

```
Qy 1 ATGTTAATGCTTTATGATTAATTTGGCCCTACTGATAGGACGATTCATCTGGCCTTA 60
Db 1 ATGATGATAATCGTCATCTACTACTAAGTTATCTTATCGGCGCTTTCCCAAGTGGATTC 60
Qy 61 ATTTGGGCAAGCTTGGCCAAAGGATTCATATTCGGGAGCACGGAAGCGGCAACTTAGGC 120
Db 61 GTAAATGGAAATATATTTTCAAAAAGATATATAGACAATTTGGTAGTGGTAATCTAGTC 120
Qy 121 GCTACCAATGCAATCCGTACATTTGGGTGTAAGAGCTGCTCGGTGCATAGCCGGAGAT 180
Db 121 GCTACTAATAGCTTTAGAGTATAGGTGCTCTCGCAGGATTCCTGGTAACATTTCTAGAT 180
Qy 181 ATTTTGAAGGACACTGGCAACTGCATGCTTTTCTCATGCAATGTTGA----- 230
Db 181 ATTTTCAAAAGGTTTCATAACTGTTTCTTCCCTTTATGTTTACCAGTTTCACGCGAGATGC 240
Qy 231 -----TATTCACCCGCTTCTGACGAGTCTTTGCGGTTTATAGGCCAC 273
Db 241 CCTATTAGTACTTTTTTACAAATGGTTAATTTGGCTTATTCGCTTACTTGGACAC 300
Qy 274 GTGTTTCCCATCTTCGCCAAATTTAAAGCGGTAAAGCGGTGGGACATCAGGAGCGTT 333
Db 301 GTTATCTCTGTTTATTTAAATTTCAAGGTGGCAAGCGGTGCAACTAGTGCAGGTGTC 360
Qy 334 TTGCTATTTTACGACCCCTGTTATTTATCAGATGGTTGCGGTATTTCTCATCTTTTA 393
Db 361 GTCTTGGGAGTCAACCCGATCTTTTACTAATCTTGAATATCTTCTTTATTGTTATG 420
Qy 394 TACTTGACTAAATTTGTTTCTCTCATCGATGTTTAAACAGGATCTACTGTTATATAT 453
Db 421 AAGATTTTAAATATGTTTCTTTAGCAAGTATCGTTGACGCAANTTGGTGTGATGGC 480
Qy 454 AGTTTCTTTGTCATGATACGTAATTTATTTGTTTACCCCTGCTCACCATTATTTTGT 513
Db 481 TCGCTTATCATCAAGACTATATTTTATTAGTCGTTAGTTTCTTAGTTCAATCATATTG 540
Qy 514 ATATACACACACCGGACGCAATTTAAACGAATTTATCAATAAACAAGACCTTAAAGTAAA 573
Db 541 ATAATTAGACATCGCTCTAATATCTCAAGAATTTTATAGAGCGGAAGAACCTTAAATAA 600
```

Qy 574 TGGTT 578
 Db 601 TGGAT 605

RESULT 9

US-09-815-242-6900
 : Sequence 6900, Application US/09815242
 : Patent No. US20020061569A1
 : GENERAL INFORMATION:
 : APPLICANT: Haselbeck, Robert
 : APPLICANT: Ohlsen, Kari L.
 : APPLICANT: Zyskind, Judith W.
 : APPLICANT: Wall, Daniel
 : APPLICANT: Tawick, John D.
 : APPLICANT: Carr, Grant J.
 : APPLICANT: Yamamoto, Robert T.
 : APPLICANT: Xu, H. Howard
 : TITLE OF INVENTION: Identification of Essential Genes in
 : TITLE OF INVENTION: Prokaryotes
 : FILE REFERENCE: ELITRA 011A
 : CURRENT APPLICATION NUMBER: US/09/815,242
 : CURRENT FILING DATE: 2001-03-21
 : PRIOR APPLICATION NUMBER: 60/191,078
 : PRIOR FILING DATE: 2000-03-21
 : PRIOR APPLICATION NUMBER: 60/206,848
 : PRIOR FILING DATE: 2000-05-23
 : PRIOR APPLICATION NUMBER: 60/207,727
 : PRIOR FILING DATE: 2000-05-26
 : PRIOR APPLICATION NUMBER: 60/242,578
 : PRIOR FILING DATE: 2000-10-23
 : PRIOR APPLICATION NUMBER: 60/253,625
 : PRIOR FILING DATE: 2000-11-27
 : PRIOR APPLICATION NUMBER: 60/257,931
 : PRIOR FILING DATE: 2000-12-22
 : PRIOR APPLICATION NUMBER: 60/269,308
 : PRIOR FILING DATE: 2001-02-16
 : NUMBER OF SEQ ID NOS: 14110
 : SOFTWARE: FastSEQ for Windows Version 4.0
 : SEQ ID NO 6900
 : LENGTH: 600
 : TYPE: DNA
 : ORGANISM: Haemophilus influenzae
 : FEATURE:
 : NAME/KEY: CDS
 : LOCATION: (1)...(600)
 : US-09-815-242-6900

Query Match 8.2%; Score 48; DB 10; Length 600;

Best Local Similarity 49.4%; Pred. No. 0.00043;

Matches 154; Conservative 0; Mismatches 155; Indels 3; Gaps 1;

Qy 12 TTTATTGATTTTGGCTACTGATAGGAGGAGTCCATCTGCTTAATTTGGGCAA 71
 Db 18 TTTTATATGCTTTTGGCTATCTTTAGGCTCAATTTCCAGTCGATTTAAATTTGTCG 77
 Qy 72 GCTTCCGAAGGATTTGATATTCGGACACGAGCGCACTAGCGCTACCAATGC 131
 Db 78 CATTCCAGGCTTCCAGATCTCTGCAAAATGGTTTACACACCCCTGGGGGCAAAATGT 137
 Qy 132 ATTCGCTACATTTGGTGTAAAGCTGGTTCGCTGCTCATAGCCGAGATATTTTGAAGG 191
 Db 138 GTTGGCATTTGGAATCTGAATCTGCTTGGCGTCTTATTTTGATATTTAAAGG 197
 Qy 192 GACACTGGC---AACTGATTTGCTTTTCTCATGATGTTGATATTCACCGCTTTTTCG 248
 Db 198 GATGATCTCTGATGGGAGGTATTTATTAGTTTAAACCCAAATTTGAATTTGGGAATGTT 257
 Qy 249 AGGAGTCTTTGGGTTTGGCCAGCTTTTCCCATCTTCCCAAAATTTAAAGGGGTAA 308
 Db 258 CGACTTGGTCTTGTGTAGGCAATATTTTCCCAATTTTCTTCAATTTAAAGGTGGCA 317
 Qy 309 AGCCGTGGCGAC 320

Db 318 AGGTGTTGCAAC 329

RESULT 10

US-09-974-300-4100
 : Sequence 4100, Application US/09974300
 : Patent No. US20020146721A1
 : GENERAL INFORMATION:
 : APPLICANT: Berka, Randy M.
 : APPLICANT: Clausen, Ib Groth
 : TITLE OF INVENTION: Methods For Monitoring Multiple Gene
 : TITLE OF INVENTION: Expression
 : FILE REFERENCE: 10085.500-US
 : CURRENT APPLICATION NUMBER: US/09/974,300
 : CURRENT FILING DATE: 2001-10-05
 : PRIOR APPLICATION NUMBER: 09/680,598
 : PRIOR FILING DATE: 2000-10-06
 : PRIOR APPLICATION NUMBER: 60/279,526
 : PRIOR FILING DATE: 2001-03-27
 : NUMBER OF SEQ ID NOS: 8481
 : SOFTWARE: FastSEQ for Windows Version 4.0
 : SEQ ID NO 4100
 : LENGTH: 72
 : TYPE: DNA
 : ORGANISM: Bacillus licheniformis
 : US-09-974-300-4100

Query Match 8.0%; Score 46.4; DB 10; Length 72;

Best Local Similarity 77.8%; Pred. No. 0.0005;

Matches 56; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy 509 TTGTGATATACAGACACCGACGACATTAACGAATTAACCAATTAATCAATAAACAACCACTTAAG 568
 Db 1 TTGTGATATACAGACACGACGACATTAACGAATTAACCAATTAATCAATAAACAACCACTTAAG 60
 Qy 569 TAAATGCTTAT 580
 Db 61 TCAAGTGATGT 72

RESULT 11

US-09-815-242-7356
 : Sequence 7356, Application US/09815242
 : Patent No. US20020061569A1
 : GENERAL INFORMATION:
 : APPLICANT: Haselbeck, Robert
 : APPLICANT: Ohlsen, Kari L.
 : APPLICANT: Zyskind, Judith W.
 : APPLICANT: Wall, Daniel
 : APPLICANT: Tawick, John D.
 : APPLICANT: Carr, Grant J.
 : APPLICANT: Yamamoto, Robert T.
 : APPLICANT: Xu, H. Howard
 : TITLE OF INVENTION: Identification of Essential Genes in
 : TITLE OF INVENTION: Prokaryotes
 : FILE REFERENCE: ELITRA 011A
 : CURRENT APPLICATION NUMBER: US/09/815,242
 : CURRENT FILING DATE: 2001-03-21
 : PRIOR APPLICATION NUMBER: 60/191,078
 : PRIOR FILING DATE: 2000-03-21
 : PRIOR APPLICATION NUMBER: 60/206,848
 : PRIOR FILING DATE: 2000-05-23
 : PRIOR APPLICATION NUMBER: 60/207,727
 : PRIOR FILING DATE: 2000-05-26
 : PRIOR APPLICATION NUMBER: 60/242,578
 : PRIOR FILING DATE: 2000-10-23
 : PRIOR APPLICATION NUMBER: 60/253,625
 : PRIOR FILING DATE: 2000-11-27
 : PRIOR APPLICATION NUMBER: 60/257,931
 : PRIOR FILING DATE: 2000-12-22
 : PRIOR APPLICATION NUMBER: 60/269,308
 : PRIOR FILING DATE: 2001-02-16

APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 7517
LENGTH: 663
TYPE: DNA
ORGANISM: Helicobacter pylori
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(663)
US-09-815-242-7517

Query Match
Best Local Similarity 55.4%; Score 39.8; DB 10; Length 663;
Matches 77; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

OY 5 TAATGCTTTATTGATATTTTGGCCCTACTTGATAGGCACCATTCATCTGGCTTAATTG 64
Db 32 TCAATGATATTTTCAACCTTTTGGCTATTTGATGGGGGATTCCTTTTGGCTATGCGT 91
OY 65 TGGCGAAGCTTGGCAAGGAATTGATTCGGGACGACGGAAGCGCAACTTAGGGGCTA 124
Db 92 TAATGAAATCTTTTATGGCATCGATATTACTAAATCGGATCGGGGGTATTGGCCCGA 151
OY 125 CCAATGCATTCGTCACATT 143
Db 152 CGAATGCTTACGGCTTT 170

RESULT 15
US-09-813-320-3
Sequence 3, Application US/09813320
Patent No. US20020142378A1
GENERAL INFORMATION:
APPLICANT: ZHANG, Hongyu et al.
TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
TITLE OF INVENTION: AND USES THEREOF
FILE REFERENCE: CL001172
CURRENT APPLICATION NUMBER: US/09/813,320
CURRENT FILING DATE: 2001-03-21
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 397658
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature

LOCATION: (1)...(397658)
OTHER INFORMATION: n = A,T,C or G
US-09-813-320-3

Query Match
Best Local Similarity 50.0%; Score 38; DB 10; Length 397658;
Matches 95; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

OY 282 CATCTTCGCCCAAAATTTAAAGCGGTAAAGCGGTGGGACACATCAGGAGCGCTTTTGGCTATT 341
Db 331668 CTTCTTATGCAATTTTCTAAGTCTCTAACTGGTTGCCTTATATCTTATGATAGTGTCTT 331727
OY 342 TTACGCACCCCTCTTATTTATACGATGGTTGCGGTATTTCTTCATCTTTTATATCTTGAC 401
Db 331728 TTACAAAGTTTAAATTTATTAACACATGCAATTTGCACTCTTCTTTCATGTTAAGTCACITGAG 331787
OY 402 TAAATTTGTTTCTCTCTCATCGATGTTAAACAGGATCTATCTGTTATATATATAGTTTCTT 461
Db 331788 GAATTTGGGTTTATTAAACTTTAGCTAAATAAACAACAGTCAGTCGCTAAGATATATTGTACA 331847
OY 462 TGTCCATGAT 471
Db 331848 AATACCTGAT 331857

Search completed: December 26, 2002, 00:37:24
Job time : 459.89 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: December 24, 2002, 22:13:55 ; Search time 11247.8 seconds
(without alignments)
1661.118 Million cell updates/sec

Title: US-10-068-080-11
Perfect score: 642
Sequence: 1 tactaatgtatcaaaataa.....tagttctaggattttttatt 642

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :	GenEmbl:*	Query				ID	Description
		Result No.	Score	Match	Length		
	1: gb.ba.*	C 1	67.2	10.5	7218	6	I66494 Sequence 14
	2: gb.htg.*	C 2	49	7.6	163034	2	AC099432 Rattus no
	3: gb.in.*	C 3	48.2	7.5	171173	2	AF481054 Mus muscu
	4: gb.cm.*	C 4	45	7.0	85786	9	AL162582 Human DNA
	5: gb.ov.*	C 5	44.6	6.9	115489	2	AC117072 Dictyoste
	6: gb.pat.*	C 6	44.4	6.9	179553	2	AC024253 Homo sapi
	7: gb.ph.*	C 7	44.4	6.9	186935	2	AC022322 Homo sapi
	8: gb.pl.*	C 8	44.4	6.9	176351	2	AC117007 Rattus no
	9: gb.pr.*	C 9	43.8	6.8	161799	2	AC098991 Rattus no
	10: gb.ro.*	C 10	43.8	6.8	211027	2	AC130161 Rattus no
	11: gb.sts.*	C 11	43.5	6.8	62352	2	AC116990 Dictyoste
	12: gb.sy.*	C 12	43.5	6.8	201097	2	AC112091 Rattus no
	13: gb.un.*	C 13	43.4	6.8	167439	2	AC102515 Mus muscu
	14: gb.vi.*	C 14	43.4	6.8	207364	2	AL671857 Mus muscu
	15: em.ba.*	C 15	43	6.7	173048	5	AL772154 zebrafish
	16: em.fun.*	C 16	42.4	6.6	196628	2	AL772388 Danio rer
	17: em.in.*	C 17	42.4	6.6	161546	2	AC108270 Rattus no
	18: em.mu.*	C 18	42	6.5	3192	3	PF837C3 Plasmodiu
	19: em.ov.*	C 19	42	6.5	3769	3	PF818C1 Plasmodiu
	20: em.pat.*	C 20	42	6.5	14955	3	AF000580 Dictyoste
	21: em.ph.*	C 21	42	6.5	79981	2	AC126315 Rattus no
	22: em.ro.*	C 22	42	6.5	197149	2	AC103297 Rattus no
	23: em.sts.*	C 23	41.5	6.5	321003	2	PF8AL4P3 Plasmodiu
	24: em.un.*	C 24	41.4	6.4	198653	2	AC114691 Rattus no
	25: em.vi.*	C 25	41.2	6.4	167364	10	AC121590 Mouse DNA
	26: em.ro.*	C 26	41.2	6.4	220275	10	AL732404 Dictyoste
	27: em.in.*	C 27	40.3	6.4	79769	2	AC115678 Rattus no
	28: em.mu.*	C 28	40.3	6.4	169163	2	AC115666 Dictyoste
	29: em.ov.*	C 29	40.5	6.3	56152	2	AC116963 Dictyoste
	30: em.pat.*	C 30	40.5	6.3	88037	2	PF8AL13P8 Plasmodiu
	31: em.ph.*	C 31	40.4	6.3	128230	9	HS313L4 Human DNA s
	32: em.sts.*	C 32	40.4	6.3	131495	2	AC120683 Rattus no
	33: em.un.*	C 33	40.4	6.3	169965	2	AC118402 Rattus no
	34: em.vi.*	C 34	40.2	6.3	5145	2	AC015315 Drosophil
	35: em.ba.*	C 35	40.2	6.3	43993	2	AC116965 Dictyoste
	36: em.fun.*	C 36	40.2	6.3	84785	2	AC117005 Rattus no
	37: em.in.*	C 37	40.2	6.3	165337	2	AC096212 Rattus no
	38: em.mu.*	C 38	40.2	6.3	182080	3	AC012373 Drosophil
	39: em.ov.*	C 39	40.2	6.3	188755	9	AC013562 Homo sapi
	40: em.pat.*	C 40	40.2	6.3	303760	3	AE003509 Drosophil
	41: em.ph.*	C 41	39.8	6.2	136431	9	AC093524 Homo sapi
	42: em.ro.*	C 42	39.8	6.2	161769	2	AC122097 Rattus no
	43: em.sts.*	C 43	39.8	6.2	194020	10	AC121808 Mus muscu
	44: em.un.*	C 44	39.8	6.2	199133	2	AC122212 Mus muscu
	45: em.vi.*	C 45	39.6	6.2	84563	5	AC096885 Danio rer

ALIGNMENTS

RESULT 1
I66494/c
LOCUS I66494
DEFINITION Sequence 14 from patent US 5670367.
ACCESSION I66494
VERSION I66494.1 GI:2724471
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 7218)
AUTHORS Dörner, F., Scheiflinger, F. and Falkner, F. Gunter.
TITLE Recombinant fowlpox virus
JOURNAL Patent: US 5670367-A 14 23-SEP-1997;
FEATURES Location/Qualifiers

Pred. No. is the number of results predicted by chance to have a

```
source 1. 7218
/organism="unknown"
BASE COUNT 1944 a 1491 c 1486 g 1929 t 368 others
ORIGIN

Query Match 10.5% Score 67.2; DB 6; Length 7218;
Best Local Similarity 5.3%; Pred. NO. 2.2e-07;
Matches 21; Conservative 226; Mismatches 149; Indels 0; Gaps 0;

QY 158 CATACCGTGGAAACACATACTGAAAGTTTCCTGGGATCGTTCGCACCAAGCGCTAAT 217
Db 1476 CCTATCTATGCAATAGTTAAAGAGATAGAGAATTTGGTACRRRRRRRRRRRRRRR 1417
QY 218 AAAAAGTAGATGTTCCGCAAGAGAGAGATGAGAAACCTGAAACCGCAATAGCGGTAT 277
Db 1416 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1357
QY 278 GGAAGGATAGAAACGCTCTAAATTTCCACATTCCTCCAGACGTTGGTCACGACCTCACT 337
Db 1356 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1297
QY 338 AAAGCCCTAAACGGGATAGAGACAGATGGAACCTATAGAGAAACCTCGACAGA 397
Db 1296 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1237
QY 398 TAGAACCGTCTATACTAAAGTCACAGATCACAGTGTCTGTAGCTAACGCGCAATAGCCCC 457
Db 1236 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1177
QY 458 AAGACGAGAAAGGTGAAACCAAAATAGGACTCATGATCTAGAGAGAGATAGCCAT 517
Db 1176 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1117
QY 518 AATAGATCGTGAACGATCAACTAATAGTAGCAG 553
Db 1116 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1081

RESULT 2
AC099432/c
LOCUS
DEFINITION Rattus norvegicus clone CH230-103G20, *** SEQUENCING IN PROGRESS
AC099432
VERSION AC099432.3 GI:21728595
KEYWORDS HTG; HTGS, PHASE1.
SOURCE Norway rat.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 163034)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amarantunga,H.C., Are,J.R., Ayele,M., Banks,T.,
Barberia,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlak,P., Hayes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hoques,M., Holloway,C., Hollins,B.,
Homsí,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jollivet,S., Joudah,S.,
Karlisson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
Nguyen,N., Nickerson,E., Nwokoko,S., Ogih,M., Okwuonu,G.,
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
Rives,M., Rojas,A., Rojokan,I., Rolfe,M., Ruiz,S., Savary,G.,
Scherer,S., Scott,G., Shen,H., Shoshitari,N., Sisson,I.,
Sodergren,E., Sonaika,T., Sparks,A., Stanley,H., Stone,H.,
Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
Tansley,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,C.,
Williams,G., Williamson,A., Wleczky,R., Woodson,S., Worley,K.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

Direct Submission
Unpublished
2 (bases 1 to 163034)
Worley,K.C.
Direct Submission
Submitted (15-NOV-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 163034)
Worley,K.C.
Direct Submission
Submitted (12-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 11, 2002 this sequence version replaced gi:17973419.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GHWE
Center clone name: CH230-103G20
----- Summary Statistics
Sequencing vector: Plasmid
Chemistry: Dye-terminator Big Dye 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 79006 bases at least Q40
Consensus quality: 84121 bases at least Q30
Consensus quality: 88465 bases at least Q20
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 82 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1382: contig of 1382 bp in length
* 1383 1482: gap of unknown length
* 1483 2616: contig of 1134 bp in length
* 2617 2716: gap of unknown length
* 2717 4049: contig of 1333 bp in length
* 4050 4149: gap of unknown length
* 4150 5644: contig of 1495 bp in length
* 5645 5744: gap of unknown length
* 5745 6913: contig of 1169 bp in length
* 6914 7014: gap of unknown length
* 7014 8388: contig of 1375 bp in length
* 8389 10138: gap of unknown length
* 10138: contig of 1650 bp in length
```

```
* 10139 10238: gap of unknown length
* 10239 11734: contig of 1496 bp in length
* 11735 11834: gap of unknown length
* 11835 13287: contig of 1453 bp in length
* 13288 13387: gap of unknown length
* 13388 14709: contig of 1322 bp in length
* 14710 14809: gap of unknown length
* 14810 15861: contig of 1052 bp in length
* 15862 15961: gap of unknown length
* 15962 17518: contig of 1557 bp in length
* 17519 17618: gap of unknown length
* 17620 19300: contig of 1682 bp in length
* 19301 19400: gap of unknown length
* 19401 20729: contig of 1329 bp in length
* 20730 20829: gap of unknown length
* 20830 21939: contig of 1100 bp in length
* 21940 22029: gap of unknown length
* 22030 23415: contig of 1386 bp in length
* 23416 23515: gap of unknown length
* 23516 25034: contig of 1519 bp in length
* 25035 25134: gap of unknown length
* 25135 26213: contig of 1079 bp in length
* 26214 26313: gap of unknown length
* 26314 27705: contig of 1392 bp in length
* 27706 27805: gap of unknown length
* 27806 29246: contig of 1441 bp in length
* 29247 29346: gap of unknown length
* 29347 30523: contig of 1177 bp in length
* 30524 30623: gap of unknown length
* 30624 32166: contig of 1543 bp in length
* 32167 32266: gap of unknown length
* 32267 33674: contig of 1408 bp in length
* 33675 33774: gap of unknown length
* 33775 35023: contig of 1249 bp in length
* 35024 35123: gap of unknown length
* 35124 37010: contig of 1887 bp in length
* 37011 37110: gap of unknown length
* 37111 38878: contig of 1768 bp in length
* 38879 38978: gap of unknown length
* 38979 40329: contig of 1351 bp in length
* 40330 40429: gap of unknown length
* 40430 41896: contig of 1467 bp in length
* 41897 41996: gap of unknown length
* 41997 43475: contig of 1479 bp in length
* 43476 43575: gap of unknown length
* 43576 44684: contig of 1109 bp in length
* 44685 44784: gap of unknown length
* 44785 46637: contig of 1853 bp in length
* 46638 46737: gap of unknown length
* 46738 47987: contig of 1250 bp in length
* 47988 48087: gap of unknown length
* 48088 49155: contig of 1068 bp in length
* 49156 49255: gap of unknown length
* 49256 50589: contig of 1334 bp in length
* 50590 50689: gap of unknown length
* 50690 51988: contig of 1299 bp in length
* 51989 52088: gap of unknown length
* 52089 53966: contig of 1878 bp in length
* 53967 54066: gap of unknown length
* 54067 55332: contig of 1266 bp in length
* 55333 55432: gap of unknown length
* 55433 56979: contig of 1547 bp in length
* 56980 57079: gap of unknown length
* 57080 58450: contig of 1371 bp in length
* 58451 58550: gap of unknown length
* 58551 61272: contig of 2722 bp in length
* 61273 61372: gap of unknown length
* 61373 62639: contig of 1267 bp in length
* 62640 62739: gap of unknown length
* 62740 64180: contig of 1441 bp in length
* 64181 64280: gap of unknown length
* 64281 65566: contig of 1286 bp in length
* 65567 65666: gap of unknown length
```

```
* 65667 66958: contig of 1292 bp in length
* 66959 67058: gap of unknown length
* 67059 68733: contig of 1675 bp in length
* 68734 68833: gap of unknown length
* 68834 70507: contig of 1674 bp in length
* 70508 70607: gap of unknown length
* 70608 71880: contig of 1273 bp in length
* 71881 71980: gap of unknown length
* 71981 73275: contig of 1295 bp in length
* 73276 73375: gap of unknown length
* 73376 74950: contig of 1575 bp in length
* 74951 75050: gap of unknown length
* 75051 76776: contig of 1726 bp in length
* 76777 76876: gap of unknown length
* 76877 77949: contig of 1073 bp in length
* 77950 78049: gap of unknown length
* 78050 79783: contig of 1734 bp in length
* 79784 79883: gap of unknown length
* 79884 82083: contig of 2200 bp in length
* 82084 82183: gap of unknown length
* 82184 83593: contig of 1410 bp in length

Query Match 7.6%; Score 49; DB 2: Length 163034;
Best Local Similarity 44.6%; Pred. No. 0.015;
Matches 193; Conservative 0; Mismatches 240; Indels 0; Gaps 0;

Qy 138 GTAATATCCATTCTTCGACCATACCGTTGGAAACACTAACTGAAAGTTCTTCTGGGA 197
Db 154783 GCAACAAACAACACAGCAACAAACAACAGCAACAGCAACAGCAACAGCAACAA 154724

Qy 198 TCGTTGGCAGGAGGCTTAATAAAGTAGATGTTCCGAAAGAGGAGAGTAGAACCTGA 257
Db 154723 GCAACAGCAACAACAACAGCAACAAACAACAGCAACAGCAACAAACAACAGCA 154664

Qy 258 AAACCGACAAATAGCCGGTATGGAAGGATAGAAACGCTCTAAATTTCCACCATTTCCGACA 317
Db 154663 ACAACAGCAACAGCAACAACAGCAACAGCAACAGCAACAGCAACAGCAACAA 154604

Qy 318 CGTTGGTCAGCCTTACTTAAAGCCTTAAACGGGATAGAACAGACAGATGGAACCTA 377
Db 154603 GCAACAGCAACAACAGCAACAAACAACAGCAACAGCAACAGCAACAGCAACAA 154544

Qy 378 ATAGAAGAAACCTCGAGAGATAGAACCGTCTATCTACTAAAGTGACAGATCACAGTCTCCTAG 437
Db 154543 ACAGCAACAACAACAACAGCAACAGCAACAGCAACAGCAACAGCAACAGCAACA 154484

Qy 438 CTAAACGCCGACATAGCCCCCAAGCAGAGAAAGGTGAAAAACCAAAATAGGACTCTTGTAT 497
Db 154483 GCAACAGCAACAACAGCAACAGCAACAAACAACAGCAACAGCAACAGCAACAA 154424

Qy 498 ACTGAGAGAGAAGTAGCGGTAATAGATCTGTGACGATCAAACTAATAGTAAGCAGTATT 557
Db 154423 GCAGCAACAGCAACAGCAACAGCAACAGCAACAGCAACAGCAACAGCAACAA 154364

Qy 558 CCTGTTATATCGA 570
Db 154363 ACGTCTTATGGA 154351
```

```
RESULT 3
LOCUS AF481054
DEFINITION Mus musculus chromosome 17 clone 201h10 strain 12956/SVEvtac, ***
SEQUENCING IN PROGRESS ***, 4 ordered pieces.
ACCESSION AF481054
VERSION AF481054.1 GI:20453846
KEYWORDS HTG: HTGS_PHASE2.
SOURCE Mus musculus.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 171173)
AUTHORS Brathwaite,M., Weeltz,P., Qian,Y., Dudekula,D., Schlessinger,D. and
```

```

Nagaraja,R.
Genomic Sequence Analysis in the Mouse t-complex Region
Unpublished
2 (bases 1 to 171173)
Brathwaite,M., Waelitz,P., Qian,Y., Dudekula,D., Schlessinger,D. and
Nagaraja,R.
Direct Submission
Submitted (05-FEB-2002) Laboratory of Genetics, National Institute
on Aging/NIH, 333 Cassell Drive, Suite 4000, Baltimore, MD 21224,
USA
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 27046: contig of 27046 bp in length
* 2 27147 64037: contig of 36891 bp in length
* 3 64038 64137: gap of unknown length
* 4 64138 139681: contig of 75544 bp in length
* 5 139682 139781: gap of unknown length
* 6 139782 171173: contig of 31392 bp in length.
FEATURES
    source
        1. .171173
        /organism="Mus musculus"
        /strain="12956/SvEvTac"
        /db_xref="taxon:10090"
        /chromosome="17"
        /clone="201h10"
        /clone_lib="RPC1-22"
        3. .43
        /rpt_family="Simple_repeat"
        /rpt_family="SINE/B4"
        complement(44..270)
        623..648
        /rpt_family="Low_complexity"
        complement(659..>23207)
        /gene="Slc22a2"
        /note="Eight exons of the Slc22a2 gene are accounted for
        with the remaining exons most likely in the gap region."
        evidence-not_experimental
        complement(join(659..1100,9260..9359,13821..13933,
        14346..14454,16549..16763,19070..19176,20713..20827,
        23039..>23207))
        /gene="Slc22a2"
        /product="organic cation transporter 2"
        /note="similar to GenBank Accession Number AJ006036"
        complement(659..664)
        /gene="Slc22a2"
        /evidence-not_experimental
        complement(join(1040..1100,9260..9359,13821..13933,
        14346..14454,16549..16763,19070..19176,20713..20827,
        23039..>23207))
        /gene="Slc22a2"
        /codon_start=3
        /evidence-not_experimental
        /product="organic cation transporter 2"
        /protein_id="AA022158.1"
        /db_xref="GI:20453848"
        /translation="TEFVGLGYRRTVCIGYQIAFTVGLLLAGVAYALPNRWLQFAV
        TLPNFCFLLYFWCIPESPRWLISONKNAKMKIHKIAKNGKSVPSVLSQSLADEPT
        GKLPNLSFLDLVTPQIKRHTLIIMNWFTSSVYQGLIMHMGLAGDNIYLDFFYSAL
        VEPFAAFTIILTDIRIGRRYPMVSNVAGAACLASVFIIDDLQWLKITVACLCRGCI
        TIAYEMVCLNNAELYPTYIRNLAVLCSSMCDIGGIIVTPELVVRLTDIWEFFLVVFA
        VGLVAGCLVLLLPETGKALPETIEDAEMQRPKKKXRIYLVQKKKALS"
        1863..2038
        /rpt_family="Simple_repeat"
        2124..2293
        /rpt_family="Simple_repeat"

```

```

repeat_region 2276..2450
                /rpt_family="Simple_repeat"
repeat_region 2504..2665
                /rpt_family="Simple_repeat"
repeat_region 2683..2857
                /rpt_family="Simple_repeat"
repeat_region 2877..3004
                /rpt_family="Simple_repeat"
repeat_region 3036..3215
                /rpt_family="Simple_repeat"
repeat_region 3255..3433
                /rpt_family="Simple_repeat"
repeat_region 3405..3581
                /rpt_family="Simple_repeat"
repeat_region 3584..3697
                /rpt_family="Simple_repeat"
repeat_region 3699..3857
                /rpt_family="Simple_repeat"
repeat_region 3865..3980
                /rpt_family="Simple_repeat"
repeat_region 3988..4170
                /rpt_family="Simple_repeat"
repeat_region 4171..4278
                /rpt_family="Simple_repeat"
repeat_region 4279..4458
                /rpt_family="Simple_repeat"
repeat_region 4460..4509
                /rpt_family="Simple_repeat"
repeat_region 4567..4745
                /rpt_family="Simple_repeat"
repeat_region 4751..4906
                /rpt_family="Simple_repeat"
repeat_region 4923..5062
                /rpt_family="Simple_repeat"
repeat_region complement(5118..5533)
                /rpt_family="LTR/ERVK"
repeat_region 5605..5782
                /rpt_family="Simple_repeat"
repeat_region 5856..6128
                /rpt_family="Simple_repeat"
repeat_region 6160..6473
                /rpt_family="Simple_repeat"
repeat_region 6475..6645
                /rpt_family="Simple_repeat"
repeat_region 6578..6757
                /rpt_family="Simple_repeat"
repeat_region 6955..7155
                /rpt_family="SINE/MIR"
                complement(8158..8341)
                /rpt_family="LINE/L1"
repeat_region complement(8652..8800)
                /rpt_family="SINE/Alu"
                8967..9032
                /rpt_family="Simple_repeat"
repeat_region 9767..9855
                /rpt_family="SINE/B4"
repeat_region 9868..9917
                /rpt_family="Simple_repeat"
repeat_region 10046..10131
                /rpt_family="SINE/Alu"
repeat_region 10147..10167
                /rpt_family="Low_complexity"
repeat_region complement(10170..10349)
                /rpt_family="SINE/B2"
                10351..10696
                /rpt_family="LTR/MaLR"
repeat_region 10697..10762
                /rpt_family="Simple_repeat"
repeat_region 10763..10780
                /rpt_family="LTR/MaLR"
repeat_region 10820..10890
                /rpt_family="SINE/ID"
repeat_region 11332..11417

```


Center code: WUGSC
 Web site: http://genome.wustl.edu/gsc/index.shtml
 ----- Project Information -----
 Center project name: H NH0317M22
 ----- Summary Statistics -----
 Sequencing vector: M13; 100%
 Sequencing vector: plasmid; 0%
 Chemistry: Dye-terminator ET; 100% of reads
 Chemistry: Dye-terminator Big Dye; 0% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 181940 bases at least Q40
 Consensus quality: 183253 bases at least Q30
 Consensus quality: 183774 bases at least Q20
 Insert size: 186000; agarose-fp
 Insert size: 185435; sum-of-contigs
 Quality coverage: 6.94 in Q20 bases; agarose-fp
 Quality coverage: 7.00 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 16 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 * 1363: contig of 1363 bp in length
 * 1364: gap of unknown length
 * 1464: contig of 1790 bp in length
 * 3253: gap of unknown length
 * 3354: contig of 1245 bp in length
 * 4598: gap of unknown length
 * 4699: contig of 1135 bp in length
 * 5833: gap of unknown length
 * 5934: contig of 2713 bp in length
 * 8646: gap of unknown length
 * 8747: contig of 3046 bp in length
 * 11792: gap of unknown length
 * 11893: contig of 2921 bp in length
 * 14814: gap of unknown length
 * 14914: contig of 4924 bp in length
 * 19837: gap of unknown length
 * 19938: contig of 4835 bp in length
 * 24772: gap of unknown length
 * 24873: contig of 6093 bp in length
 * 30966: gap of unknown length
 * 31066: contig of 9776 bp in length
 * 40842: gap of unknown length
 * 40842: contig of 11928 bp in length
 * 52869: gap of unknown length
 * 52870: contig of 10666 bp in length
 * 63636: gap of unknown length
 * 63736: contig of 26487 bp in length
 * 90222: gap of unknown length
 * 90223: contig of 45599 bp in length
 * 90323: gap of unknown length
 * 135922: contig of 50914 bp in length.
 * 136022: 186935: contig of 50914 bp in length.
 *
 * Location/Qualifiers
 * 1..186935
 * /organism="Homo sapiens"
 * /db_xref="taxon:9606"
 * /chromosome="6"
 * /clone="RP11-317M22"
 * 1..1363
 * /note="assembly_name:Contig7"
 * 1464..3253
 * /note="assembly_name:Contig8"
 * 3354..4598
 * /note="assembly_name:Contig9"
 * 4699..5833
 * /note="assembly_name:Contig10"
 * 5934..8646
 * /note="assembly_name:Contig11"

FEATURES
source

misc_feature 8747..11792
 /note="assembly_name:Contig12"
 misc_feature 11893..14813
 /note="assembly_name:Contig13"
 misc_feature 14914..19837
 /note="assembly_name:Contig14"
 misc_feature 19938..24772
 /note="assembly_name:Contig15"
 misc_feature 24873..30965
 /note="assembly_name:Contig16"
 misc_feature 31066..40841
 /note="assembly_name:Contig17"
 misc_feature 40942..52869
 /note="assembly_name:Contig18"
 misc_feature 52970..63635
 /note="assembly_name:Contig19"
 misc_feature 63736..90222
 /note="assembly_name:Contig20"
 clone_end:77
 vector_side:right
 misc_feature 90323..135921
 /note="assembly_name:Contig21"
 misc_feature 136022..186935
 /note="assembly_name:Contig22"
 clone_end:SP6
 vector_side:left
 BASE COUNT 53462 a 40007 c 40377 g 51585 t 1504 others
 ORIGIN
 Query Match 6.9%; Score 44.4; DB 2; Length 186935;
 Best Local Similarity 46.2%; Pred. No. 0.26;
 Matches 147; Conservative 0; Mismatches 171; Indels 0; Gaps 0;
 QY 215 AATAAAGTAGATGTTCCGAAGAGGAGTAGAAGCTGAAACCGGACCAATAGCCGG 274
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 7941 AACACAGCAACAGCAGCAGCAACAATAAGAGCAGCAGCAGCAGCAGCAGCAGCAGC 8000
 QY 275 TATGGAAGGATAGAAACGCTCTAAATTTCCACATTTCCGACAGCGTTGGTCACGACCTC 334
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 8001 AGCAGCAACACAGCAGCAACAATAAGCAACGATAGCAACAGTATCAGCAACAGCAAC 8060
 QY 335 ACTAAAGCCTAAACCGGATAGACAGAGATGGAAGCTTAATAGAGAAACCTCGAG 394
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 8061 AGCNTCAGCAACACAGCAGCAACAATAGCAGCAGTAAACAGCAGCAGCAGCAGCAAC 8120
 QY 395 AGATAGAACCGTCATCTACTAAAGTGACAGATCACAGTGTGCTAGCTAACGCCGCAATAGC 454
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 8121 AGCAGCAGCAGCAGCAACAATAAGAGCAGCAGCAACAGCAACAGTAGCAGCAGCAAC 8180
 QY 455 CCCAGACGAGAAAGGTGAAAAACCAAAATAGGACTCATTTGATCTAGAGAGAAAGTAGC 514
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 8181 AACAGCAGCAACAAACAAAGCAACAATAGCAACAGTATCAGCAACAGCAGCAACATAGC 8240
 QY 515 GATATAGAAATCGTGAAC 532
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 Db 8241 AACAAAGGATCAGCAAC 8258
 RESULT 8
 AC117007/c 176351 bp DNA linear HTG 17-JUL-2002
 LOCUS Rattus norvegicus clone CH230-358N16, *** SEQUENCING IN PROGRESS
 DEFINITION *** 61 unordered pieces.
 ACCESSION AC117007
 VERSION AC117007.4 GI:21746066
 KEYWORDS HTG; HTGS_PHASE1.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 REFERENCE 1 (bases 1 to 176351)
 AUTHORS Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,

Alsbrooks, S.L., Amaratunge, H.C., Are, J.R., Avele, M., Banks, T.,
 Barbara, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D.,
 Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P.,
 Burch, C., Burch, P., Burdett, C., Burrell, K.L., Byrd, N.C.,
 Caron, T., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
 Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
 Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,
 Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
 Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
 Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
 Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
 Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
 Garrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
 Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
 Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B.,
 Homsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
 Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
 Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseged, H.,
 Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
 Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
 Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
 Nguyen, N., Nickerson, E., Nwokenwo, S., Ogun, M., Okunolu, G.,
 Otagunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,
 Rives, M., Rojas, A., Rojokan, I., Rolfe, M., Ruiz, S., Savary, G.,
 Scher, S., Scott, G., Shen, H., Shoostari, N., Sisson, I.,
 Sodergren, E., Sonake, T., Sparks, A., Stanley, H., Stone, H.,
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
 Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,
 Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G., and Gibbs, R.

Direct Submission
 Unpublished
 2 (bases 1 to 176351)
 Worley, K.C.

Direct Submission
 Submitted (06-APR-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 176351)
 Worley, K.C.

Direct Submission
 Submitted (17-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 14, 2002 this sequence version replaced gi:20136787.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GTFI
 Center clone name: CH230-358N16
 ----- Summary Statistics
 Sequencing vector: Plasmid;
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990329
 Consensus quality: 120888 bases at least Q40
 Consensus quality: 127504 bases at least Q30
 Consensus quality: 132322 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently

* consists of 61 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1019: contig of 1019 bp in length
 " 1020 1119: gap of unknown length
 " 1120 2779: contig of 1660 bp in length
 " 2780 2879: gap of unknown length
 " 2880 4496: contig of 1617 bp in length
 " 4497 4596: gap of unknown length
 " 4597 5621: contig of 1025 bp in length
 " 5622 5721: gap of unknown length
 " 5722 6889: contig of 1168 bp in length
 " 6890 8124: contig of 1135 bp in length
 " 8125 8224: gap of unknown length
 " 8225 9804: contig of 1580 bp in length
 " 9805 9904: gap of unknown length
 " 9905 11392: contig of 1488 bp in length
 " 11393 11492: gap of unknown length
 " 11493 12676: contig of 1184 bp in length
 " 12677 12776: gap of unknown length
 " 12777 13861: contig of 1085 bp in length
 " 13862 13961: gap of unknown length
 " 13962 15707: contig of 1746 bp in length
 " 15708 15807: gap of unknown length
 " 15808 17582: contig of 1775 bp in length
 " 17583 17682: gap of unknown length
 " 17683 19005: contig of 1323 bp in length
 " 19006 19105: gap of unknown length
 " 19106 21071: contig of 1966 bp in length
 " 21072 21171: gap of unknown length
 " 21172 22855: contig of 1684 bp in length
 " 22856 22955: gap of unknown length
 " 22956 24233: contig of 1468 bp in length
 " 24234 24523: gap of unknown length
 " 24524 26272: contig of 1749 bp in length
 " 26273 26372: gap of unknown length
 " 26373 28549: contig of 2177 bp in length
 " 28550 28649: gap of unknown length
 " 28650 30139: contig of 1490 bp in length
 " 30140 30239: gap of unknown length
 " 30240 31817: contig of 1578 bp in length
 " 31818 33762: contig of 1845 bp in length
 " 33763 33862: gap of unknown length
 " 33863 36122: contig of 2260 bp in length
 " 36123 36222: gap of unknown length
 " 36223 38477: contig of 2255 bp in length
 " 38478 38577: gap of unknown length
 " 38578 40590: contig of 2013 bp in length
 " 40591 43154: contig of 2464 bp in length
 " 43155 43254: gap of unknown length
 " 43255 45512: contig of 2258 bp in length
 " 45513 47325: contig of 1713 bp in length
 " 47326 47425: gap of unknown length
 " 47426 49849: contig of 2424 bp in length
 " 49850 49949: gap of unknown length
 " 49950 51626: contig of 1677 bp in length
 " 51627 51726: gap of unknown length
 " 51727 52857: contig of 1131 bp in length
 " 52858 52957: gap of unknown length
 " 52958 56168: contig of 3211 bp in length
 " 56169 56268: gap of unknown length
 " 56269 59010: contig of 2742 bp in length
 " 59011 61204: contig of 2094 bp in length
 " 61205 61304: gap of unknown length

LOCUS	AC098991	161799 bp	DNA	linear	HTG 12-JUL-2000
DEFINITION	Rattus norvegicus clone CH230-84N8, *** SEQUENCING IN PROGRESS ***, 71 unordered pieces.				
ACCESSION	AC098991				
VERSION	AC098991.4 GI:21728526				
KEYWORDS	HTG: HTGS_PHASE1				
SOURCE	Norway rat				
ORGANISM	Rattus norvegicus				
REFERENCE	1 (bases 1 to 161799)				
AUTHORS	Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C., Alsbrooks, S.L., Amaral, H.C., Are, J.R., Ayale, M., Banks, T., Barabara, J., Benton, J., Bimaye, K., Blankenburg, K., Bonnin, D., Eouach, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Fuhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.M., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, P., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hoques, M., Holloway, C., Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseg, H., Lazo, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokwenkwo, S., Ogum, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Patters, L., Picken, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojubokan, I., Rolfe, M., Ruiz, S., Savory, G., Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I., Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabori, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansy, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Woodson, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.				
TITLE	Direct Submission				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 161799)				
AUTHORS	Worley, K.C.				
TITLE	Direct Submission				
JOURNAL	Submitted (08-NOV-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA				
REFERENCE	3 (bases 1 to 161799)				
AUTHORS	Worley, K.C.				
TITLE	Direct Submission				
JOURNAL	Submitted (12-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA				
COMMENT	On Jul 11, 2002 this sequence version replaced gi:17973510.				

RESULT 9
AC098991

Center project name: GH2A
Center clone name: CH230-84N8
----- Summary Statistics -----
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 98944 bases at least Q40
Consensus quality: 103846 bases at least Q30
Consensus quality: 107298 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 71 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```
1 1026: contig of 1026 bp in length
* 1027 1126: gap of unknown length
* 1127 2283: contig of 1157 bp in length
* 2284 2383: gap of unknown length
* 2384 3841: contig of 1458 bp in length
* 3842 3941: gap of unknown length
* 3942 5001: contig of 1060 bp in length
* 5002 5101: gap of unknown length
* 5102 6380: contig of 1279 bp in length
* 6381 6480: gap of unknown length
* 6481 7520: contig of 1040 bp in length
* 7521 7620: gap of unknown length
* 7621 9120: contig of 1500 bp in length
* 9121 9220: gap of unknown length
* 9221 11106: contig of 1886 bp in length
* 11107 12432: contig of 1226 bp in length
* 12433 12532: gap of unknown length
* 12533 14227: contig of 1695 bp in length
* 14228 14327: gap of unknown length
* 14328 15432: contig of 1105 bp in length
* 15433 15532: gap of unknown length
* 15533 16801: contig of 1269 bp in length
* 16802 16901: gap of unknown length
* 16902 18301: contig of 1400 bp in length
* 18302 18401: gap of unknown length
* 18402 20125: contig of 1724 bp in length
* 20126 21385: gap of unknown length
* 20226 21385: contig of 1160 bp in length
* 21386 21485: gap of unknown length
* 21486 22912: contig of 1427 bp in length
* 22913 23012: gap of unknown length
* 23013 24420: contig of 1408 bp in length
* 24421 24520: gap of unknown length
* 24521 25789: contig of 1269 bp in length
* 25790 25889: gap of unknown length
* 25890 27054: contig of 1165 bp in length
* 27055 27154: gap of unknown length
* 27155 28782: contig of 1628 bp in length
* 28783 28882: gap of unknown length
* 28883 30597: contig of 1715 bp in length
* 30598 30697: gap of unknown length
* 30698 32380: contig of 1683 bp in length
* 32381 32480: gap of unknown length
* 32481 33565: contig of 1085 bp in length
* 33566 33665: gap of unknown length
* 33666 35037: contig of 1372 bp in length
* 35038 35137: gap of unknown length
* 35138 36211: contig of 1074 bp in length
* 36212 36311: gap of unknown length
* 36312 37755: contig of 1444 bp in length
* 37756 37855: gap of unknown length
* 37856 40060: contig of 2205 bp in length

40061 40160: gap of unknown length
* 40161 41319: contig of 1159 bp in length
* 41320 41419: gap of unknown length
* 41420 42572: contig of 1153 bp in length
* 42573 42672: gap of unknown length
* 42673 43734: contig of 1062 bp in length
* 43735 43834: gap of unknown length
* 43835 45033: contig of 1199 bp in length
* 45034 45133: gap of unknown length
* 45134 46717: contig of 1584 bp in length
* 46718 46817: gap of unknown length
* 46818 48748: contig of 1931 bp in length
* 48749 48848: gap of unknown length
* 48849 49851: contig of 1003 bp in length
* 49852 49951: gap of unknown length
* 49952 51556: contig of 1605 bp in length
* 51557 51656: gap of unknown length
* 51657 53589: contig of 1933 bp in length
* 53590 53689: gap of unknown length
* 53690 54840: contig of 1151 bp in length
* 54841 54940: gap of unknown length
* 54941 56566: contig of 2026 bp in length
* 56567 57066: gap of unknown length
* 57067 59250: contig of 2184 bp in length
* 59251 59350: gap of unknown length
* 59351 60626: contig of 1276 bp in length
* 60627 60726: gap of unknown length
* 60727 63021: contig of 2295 bp in length
* 63022 63121: gap of unknown length
* 63122 65635: contig of 2514 bp in length
* 65636 65735: gap of unknown length
* 65736 67335: contig of 1600 bp in length
* 67336 67435: gap of unknown length
* 67436 69865: contig of 2430 bp in length
* 69866 71294: contig of 1329 bp in length
* 71295 71394: gap of unknown length
* 71395 72632: contig of 1238 bp in length
* 72633 72732: gap of unknown length
* 72733 75636: contig of 2904 bp in length
* 75637 75736: gap of unknown length
* 75737 77411: contig of 1675 bp in length
* 77412 77511: gap of unknown length
* 77512 80265: contig of 2754 bp in length
* 80266 80365: gap of unknown length
* 80366 82678: contig of 2313 bp in length
* 82679 82778: gap of unknown length
* 82779 84818: contig of 2040 bp in length
* 84819 84918: gap of unknown length
* 84919 87077: contig of 2159 bp in length
* 87078 87177: gap of unknown length
* 87178 89059: contig of 1882 bp in length
* 89060 91379: contig of 2220 bp in length

Query Match 6.8%; Score 43.8; DB 2; Length 161799;
Best Local Similarity 45.0%; Pred No. 0.38;
Matches 165; Conservative 0; Mismatches 202; Indels 0; Gaps 0;

QY 209 AAGCGCTAATAAAAGTAGATGTTCCGCAAGAGGAGGAGTAGAAGCTGAACCTGAACACCGACAAT 268
DB 87372 AACAAGCAGGAGGAAGGAGATGAACAAGAACAGAACAGAACAGAACAGAACAGAACAGC 87431
QY 269 AGCCGGTATGGAAGGATAGAAGCTGCTTAATTTCCACCATTTCCGACAGCGTGGTGCAC 328
DB 87432 AACAACAAGCAGCAGCAACGACCAACAACAACAACAACAACAACAACAACAACAACAAC 87491
QY 329 GACCTCACTAAAGCCCTAAACCGGATAGAGACAGATGGAACCTTAATAGAACAAC 388
DB 87492 AACAAGCAGCAACAGCAACCAACAACAACAACAACAACAACAACAACAACAACAACAGC 87551
QY 389 CTGAGAGATAGAACCGTCATCTACTAAAGTGACAGATGACAGTGTCTGTAGTAAACGCCGAC 448
```


Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Mahoney,E., Mawhinney,E., McLeod,M.P., Meador,M., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokwkw,S., Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I., Sodergren,E., Sonaite,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Teifrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera-V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G., and Gibbs,R.

Direct Submission
Unpublished
2 (bases 1 to 201097)
Worley K.C.

Direct Submission
Submitted (19-FEB-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 201097)
Worley K.C.

Direct Submission
Submitted (13-JUL-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Jul 13, 2002 this sequence version replaced gi:19111309.

----- Genome Center -----
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information -----
Center project name: GRWY
Center clone name: CH230-50E6
----- Summary Statistics -----
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 157842 bases at least Q40
Consensus quality: 163737 bases at least Q30
Consensus quality: 169262 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
* ([see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 73 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

1 1004: contig of 1004 bp in length
* 1005 1104: gap of unknown length
* 1105 2146: contig of 1042 bp in length
* 2147 2246: gap of unknown length
* 2247 3645: contig of 1399 bp in length
* 3646 3745: gap of unknown length
* 3746 4955: contig of 1210 bp in length
* 4956 5055: gap of unknown length
* 5056 6381: contig of 1326 bp in length
* 6382 6481: gap of unknown length
* 6482 7690: contig of 1209 bp in length
* 7691 7790: gap of unknown length
* 7791 8876: contig of 1086 bp in length
* 8877 8976: gap of unknown length
*

MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheeters, R., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE

Direct Submission
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE

AUTHORS

3 (bases 1 to 167439)

Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A., Cook, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B., Harton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C., McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schupback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE

JOURNAL

COMMENT

Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 21, 2002 this sequence version replaced gi:17061601.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L18947

Center clone name: 425_G_1

----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 156127 bases at least Q40

Consensus quality: 161465 bases at least Q30

Consensus quality: 163205 bases at least Q20

Insert size: 180000; agarose-fp

Insert size: 164039; sum-of-ctotigs

Quality coverage: 4.2 in Q20 bases; agarose-fp

Quality coverage: 4.6 in Q20 bases; sum-of-ctotigs

* NOTE: This is a 'working draft' sequence. It currently consists of 35 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

* 1 699: contig of 699 bp in length

* 700 799: gap of 100 bp

* 800 2116: contig of 1317 bp in length

* 2117 2216: gap of 100 bp

* 2217 3240: contig of 1024 bp in length
* 3241 3340: gap of 100 bp
* 3341 4407: contig of 1067 bp in length
* 4408 4507: gap of 100 bp
* 4508 5591: contig of 1084 bp in length
* 5592 5691: gap of 100 bp
* 5692 6754: contig of 1063 bp in length
* 6755 6854: gap of 100 bp
* 6855 7975: contig of 1121 bp in length
* 7976 8075: gap of 100 bp
* 8076 9470: contig of 1395 bp in length
* 9471 9570: gap of 100 bp
* 9571 10856: contig of 1286 bp in length
* 10857 10956: gap of 100 bp
* 10957 12838: contig of 1882 bp in length
* 12839 12938: gap of 100 bp
* 12939 14076: contig of 1138 bp in length
* 14077 14176: gap of 100 bp
* 14177 15770: contig of 1594 bp in length
* 15771 15870: gap of 100 bp
* 15871 17857: contig of 1987 bp in length
* 17858 17957: gap of 100 bp
* 17958 20340: contig of 2383 bp in length
* 20341 20440: gap of 100 bp
* 20441 22571: contig of 2131 bp in length
* 22572 22671: gap of 100 bp
* 22672 24605: contig of 1934 bp in length
* 24606 24705: gap of 100 bp
* 24706 27186: contig of 2481 bp in length
* 27187 27286: gap of 100 bp
* 27287 29635: contig of 2349 bp in length
* 29636 29735: gap of 100 bp
* 29736 31986: contig of 2251 bp in length
* 31987 32086: gap of 100 bp
* 32087 34364: contig of 2278 bp in length
* 34365 34464: gap of 100 bp
* 34465 37780: contig of 3316 bp in length
* 37781 37880: gap of 100 bp
* 37881 40408: contig of 2528 bp in length
* 40409 40508: gap of 100 bp
* 40509 44735: contig of 4227 bp in length
* 44736 44835: gap of 100 bp
* 44836 47796: contig of 2961 bp in length
* 47797 47896: gap of 100 bp
* 47897 54045: contig of 6149 bp in length
* 54046 54145: gap of 100 bp
* 54146 60050: contig of 5905 bp in length
* 60051 60150: gap of 100 bp
* 60151 65491: contig of 5341 bp in length
* 65492 65591: gap of 100 bp
* 65592 74597: contig of 9006 bp in length
* 74598 74697: gap of 100 bp
* 74698 82538: contig of 7841 bp in length
* 82539 82638: gap of 100 bp
* 82639 91410: contig of 8772 bp in length
* 91411 91510: gap of 100 bp
* 91511 101578: contig of 10068 bp in length
* 101579 101678: gap of 100 bp
* 101679 111219: contig of 9541 bp in length
* 111220 111319: gap of 100 bp
* 111320 123455: contig of 12136 bp in length
* 123456 123555: gap of 100 bp
* 123556 135268: contig of 11713 bp in length
* 135269 135368: gap of 100 bp
* 135369 167439: contig of 32071 bp in length.

FEATURES

source

1. 167439

/organism="Mus musculus"

/db_xref="taxon:10090"

/clone_lib="RP24-425G1"

/misc_feature 1. 699

/note="assembly_fragment"

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 24, 2002, 19:31:45 ; Search time 150.797 Seconds
(without alignments)
9587.638 Million cell updates/sec

Title: US-10-068-080-11

Perfect score: 642

Sequence: 1 tactaatgtatcaaaataa.....tagttctaggattttttatt 642

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002:*

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:*
- 24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	40.2	6.3	5688	23 ABL26543	Drosophila melanog
2	40.2	6.3	7588	23 ABL26542	Drosophila melanog
3	38.4	6.0	7337	23 ABL30176	Drosophila melanog
4	38.2	6.0	5613	15 AAO63241	Crucifer 1-aminoc
5	38.2	6.0	7034	22 ABA14938	Human nervous syst
6	38.2	6.0	7036	22 ABA14937	Human nervous syst
7	38	5.9	727	24 ABQ29312	Oligonucleotide fo
8	38	5.9	727	24 ABQ29313	Oligonucleotide fo
9	36.8	5.7	5139	21 AAA70139	Plasmodium falci

C 10	36.6	5.7	7802	22 AAS45351	Chemically pretrea
C 11	36.6	5.7	7802	24 ABL32765	Human immune syste
C 12	36.6	5.7	7802	24 ABK28184	DNA transcription
C 13	36.4	5.7	807	22 AAL21312	Human breast cance
C 14	36.4	5.7	15428	23 ABL02692	Drosophila melanog
C 15	36.2	5.6	8946	24 ABL32911	Human immune syste
C 16	36	5.6	575	24 ABQ51962	Oligonucleotide fo
C 17	36	5.6	575	24 ABQ51963	Oligonucleotide fo
C 18	36	5.6	15046	24 ABL34101	Human immune syste
C 19	35.8	5.6	5394	24 ABL33008	Human immune syste
C 20	35.6	5.5	640681	24 ABA92787	Buchnera sp. genom
C 21	35.4	5.5	8905	20 AAX20263	Borrelia burgdorfe
C 22	35.2	5.5	534	24 ABQ33814	Oligonucleotide fo
C 23	35.2	5.5	534	24 ABQ33815	Oligonucleotide fo
C 24	34.8	5.4	5935	22 AAS45427	Chemically pretrea
C 25	34.8	5.4	6334	24 ABL33212	Human immune syste
C 26	34.4	5.4	362	22 AAH57352	Human brain specif
C 27	34.4	5.4	803	24 ABQ44790	Oligonucleotide fo
C 28	34.4	5.4	803	24 ABQ44791	Oligonucleotide fo
C 29	34.4	5.4	2054	24 ABK34846	Human cDNA for nov
C 30	34.4	5.4	2306	22 AAK94865	Human full-length
C 31	34.4	5.4	2448	22 AAH57560	Human brain cell s
C 32	34.2	5.3	559	24 ABQ21504	Oligonucleotide fo
C 33	34.2	5.3	559	24 ABQ21505	Oligonucleotide fo
C 34	34.2	5.3	636	23 AAS92306	DNA encoding novel
C 35	34.2	5.3	646	24 ABQ46952	Oligonucleotide fo
C 36	34.2	5.3	646	24 ABQ46953	Oligonucleotide fo
C 37	34.2	5.3	744	23 AAS77979	DNA encoding novel
C 38	34.2	5.3	4590	22 AAH24065	Yeast AOD9604-asso
C 39	34	5.3	694	24 ABQ19008	Oligonucleotide fo
C 40	34	5.3	694	24 ABQ19009	Oligonucleotide fo
C 41	33.8	5.3	9774	22 ABA19485	Human nervous syst
C 42	33.6	5.2	521	24 ABQ37468	Oligonucleotide fo
C 43	33.6	5.2	521	24 ABQ37469	Oligonucleotide fo
C 44	33.6	5.2	664	22 AAK63846	Human immune/haem
C 45	33.6	5.2	5059	20 AAX84332	Stealth virus nucl

ALIGNMENTS

RESULT 1

ABL26543
ID ABL26543 standard; DNA: 5688 BP.

XX ABL26543;

AC ABL26543;

XX 26-MAR-2002 (first entry)

DT Drosophila melanogaster genomic polynucleotide SEQ ID NO 31102.

DE Drosophila; developmental biology; cell signalling; insecticide;

XX Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical; gene; ds.

KW pharmaceutical; gene; ds.

XX Drosophila melanogaster.

OS Drosophila melanogaster.

PN WO200171042-A2.

XX 27-SEP-2001.

PD 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

DR New isolated nucleic acid detection reagent for detecting 1000 or more

PT genes from Drosophila and for elucidating cell signalling and cell-cell

DT 23-JAN-2002 (first entry)
XX Human nervous system related polynucleotide SEQ ID NO 7269.
DE
XX Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; anti-inflammatory; anti-HIV; antibacterial; vulnery;
KW antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; anticancer; anticonvulsant; antifungal;
KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
XX Homo sapiens.
XX WO200159063-A2.
XX
XX 16-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01334.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230437.
PR 08-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.

	05-DEC-2000; 2000US-0251988.	
PPR	05-DEC-2000; 2000US-0256719.	
PPR	06-DEC-2000; 2000US-0251479.	
PPR	08-DEC-2000; 2000US-0251856.	
PPR	08-DEC-2000; 2000US-0251868.	
PPR	08-DEC-2000; 2000US-0251869.	
PPR	08-DEC-2000; 2000US-0251989.	
PPR	08-DEC-2000; 2000US-0251990.	
PPR	11-DEC-2000; 2000US-0254097.	
PPR	05-JAN-2001; 2001US-0259678.	
XXX	(HUMA-) HUMAN GENOME SCI INC.	
XX	Rosen CA, Barash SC, Ruben SM;	
PFI	WPI: 2001-541565/60.	
XXX	Nucleic acids encoding 3224 human nervous system antigen polypeptides,	
PPT	useful for preventing, diagnosing and/or treating nervous system	
PPT	cancers and metastases -	
XXX	Disclosure: SEQ ID NO 7269; 1701pp + Sequence Listing; English.	
PPS	The invention relates to novel genes (ABAI1004-ABA21534) and proteins	
CCC	(ABB14678-ABB18001) useful for preventing, treating or ameliorating	
CCC	medical conditions e.g. by protein or gene therapy. The genes are	
CCC	isolated from a range of human tissues disclosed in the specification.	
CCC	The nucleic acids, proteins, antibodies and (ant)agonists are useful	
CCC	in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast	
CCC	and ovarian cancer and other cancers of the adrenal gland, bone, bone	
CCC	marrow, breast, gastrointestinal tract, liver, lung, or urogenital;	
CCC	(b) immune disorders e.g. Addison's disease, allergies, autoimmune	
CCC	haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's	
CCC	disease, multiple sclerosis, rheumatoid arthritis and ulcerative	
CCC	colitis; (c) cardiovascular disorders such as myocardial ischaemias;	
CCC	(d) wound healing; (e) neurological diseases e.g. cerebral anoxia and	
CCC	epilepsy; and (f) infectious diseases such as viral, bacterial, fungal	
CCC	and parasitic infections.	
CCC	Note: The sequence data for this patent did not form part of the	
CCC	printed specification, but was obtained in electronic format directly	
CCC	from WIPO at ftp.wipo.int/pub/published_pct_sequences.	
XX	Sequence 7034 BP; 2150 A; 1284 C; 1367 G; 2233 T; 0 other;	
SQ		
	Query Match 6.0%; Score 38.2; DB 22; Length 7034;	
	Best Local Similarity 44.28; Pred. No. 0.68;	
	Matches 157; Conservative 0; Mismatches 198; Indels 0; Gaps	
QY	140 AAAATCCATTCTTTTCGACCACATACCGTGTGGAAACAACACTAAGTGAAAAGTTTTCTTCTGGGATC 199	
Ddb		
	4014 AAAATGATAATTATAGAATAATTATATGAAAAAGGAAGTTAACAAAAGACAGAGGATAGA 3955	
QY	200 GTTGGCAGNAGGCCTAATAAAAAGTAGATGTTCGCCAAGAGGAGAGTAGNAACCTGAAA 259	
Ddb		
	3954 TTAAACAAGGTCAAACATAGCTTAATAAGGGTCTCTGGAAGAGAGAANAATTGAGAAGGTGGGA 3895	
QY	260 ACCGCAATATAGCCGTATGGAAGGCGATGAACACGCTCTAAATTTCCACCATTTCCGCACAGC 319	
Ddb		
	3894 AGATGNATAGACAAAAGATGATGAAAAAATAATTCACGAATTTGTAAGATGTAAGTTG 3835	
QY	320 GTTGTGTCAGCCTCACATAAAGCCGATTAACCGCGGATAGAAGACAGAGATGGAACGCTAAT 379	
Ddb		
	3834 TTAGATCAATCATCAAACAAAAGTTTGAAGTTAAATAAAGGAAAAGACATATCAGGAAG 3775	
QY	380 AGAAGNAACTCGAGATAGAACCGCTCATACTAAGTGACAGATACACAGTGTGCTAGCT 439	
Ddb		
	3774 TTACTAATCAAAAGGAAGATGGTAGCACAATTAATAATCAGAAAAATAGAGTTTCAGGT 3715	
QY	440 AACGCCGCAATAGCCCCAAGACGAGAGAGTGAAGAAAACCAAAATAGGACTCATT 494	
Ddb		
	3714 AAAAGCATTTATTACAATGAAAAATAGGCATTAATTAATGATAAAGGATCAATT 3660	

PD XX 06-DEC-2001.
 PF XX
 XX 06-APR-2001; 2001WO-EP03973.
 PR XX
 PR 06-APR-2000; 2000DE-1019058.
 PR 07-APR-2000; 2000DE-1019173.
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 PA (EPIG-) EPIGENOMICS AG.
 XX
 XX Olek A, Piepenbrock C, Berlin K;
 PI
 XX WPI; 2002-090046/12;
 DR
 XX
 XX New nucleic acids or oligomers, useful for diagnosing or treating
 PT diseases associated with DNA transcription, e.g. immunological
 PT disorders, Werner syndrome, psoriasis, myocardial infarction, solid
 PT tumours or cancer
 XX
 PS Claim 1: SEQ ID NO 58; 32pp; English.
 XX
 CC The invention relates to a nucleic acid, which comprises a segment of the
 CC chemically pretreated DNA of genes associated with DNA transcription from
 CC one of 346 sequences, and an oligomer, in particular an oligonucleotide
 CC or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical
 CC to the chemically pretreated DNA of genes associated with DNA
 CC transcription. The set of oligomer probes are useful for detecting the
 CC cytosine methylation state and/or single nucleotide polymorphisms (SNPs)
 CC in a chemically pretreated genomic DNA. The nucleic acids are useful for
 CC diagnosing or treating diseases associated with DNA transcription
 CC (particularly with the methylation status), e.g. adenine deaminase
 CC deficiency, viral infection, retroviral infection, Sezary syndrome,
 CC haematological disorders, immunological disorders, Werner syndrome,
 CC tuberculous, developmental disorders, psoriasis, Rieger's syndrome,
 CC neurological disorders, neurodegenerative disorders, Waardenburg
 CC syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial
 CC infarction, hypertension, angiodysplasia, erythropoiesis, congenital heart
 CC disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours
 CC or cancer. Sequences ABK28127-ABK28472 represent DNA transcription
 CC associated genomic DNA molecules of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification but was obtained in electronic format directly from the
 CC European Patent Office.
 XX
 SQ Sequence 7802 BP; 1718 A; 478 C; 2409 G; 3197 T; 0 other;
 Query Match 5.7%; Score 36.6; DB 24; Length 7802;
 Best Local Similarity 49.2%; Pred. No. 2.1;
 Matches 96; Conservative 0; Mismatches 99; Indels 0; Gaps 0;
 QY 221 AAGTAGATGTTCCGCAAGAGGAGTAGAAGACCTGAAACCCGATATGGA 280
 DB 4669 AATATACCATTTAGCAAAACGTAACCTTAAACCCGCAACCCGCAAAATCCCGCTAA 4610
 QY 281 AGGATAGAACGTCCTTAAATTTCCACCATTTCCGACAGCGTTTGGTCCACGACCTCACTAAA 340
 DB 4609 ATCCCGACCAAAACCGAATTTAAATTCGATACAAACCCGAAACTAATTTCTTAACTAT 4550
 QY 341 AGCCTTAACCCGGATAGACAGAGATGGAACGCTTAATAGAGAACCTTCGAGAGATAG 400
 DB 4549 AAAAAAACCCGCAAAACCGAATTAATTAATTAATTCGCAACGAAACGAAAAAATTA 4490
 QY 401 AACCGTCATCTACTAAA 415
 DB 4489 ATCCCTAATAATAAA 4475
 RESULT 13
 AAL21312/c
 ID AAL21312 standard; cDNA; 807 BP.
 XX
 AC AAL21312;

XX 07-DEC-2001 (first entry)
 DT
 XX Human breast cancer expressed polynucleotide 13769.
 DE
 XX Human; breast cancer; cell marker; cytostatic; ss.
 KW
 XX Homo sapiens.
 OS
 XX WO200151628-A2.
 PN
 XX 19-JUL-2001.
 PD
 XX 10-JAN-2001; 2001WO-US00798.
 PF
 XX 14-JAN-2000; 2000US-0176077.
 PR 14-MAR-2000; 2000US-0189167.
 PR 24-MAR-2000; 2000US-0192099.
 PR 29-MAR-2000; 2000US-0193480.
 PR 15-MAY-2000; 2000US-0205230.
 PR 09-JUN-2000; 2000US-0211315.
 PR 23-JUL-2000; 2000US-0220534.
 XX
 (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 PA
 XX Lillie J, Xu Y, Wang Y, Steinmann K;
 PI
 XX WPI; 2001-451856/48.
 DR
 XX New peptide useful as a marker for the diagnosis of breast cancer
 PT
 XX Claim 1: Page 2449; 3695pp; English.
 PS
 CC The invention relates to human breast cancer expressed polynucleotides
 CC (AAL07544-AAL26789) and methods of assessing whether a patient is
 CC afflicted with breast cancer by examining the correlation between the
 CC expression of certain markers and the cancerous state of breast cells.
 CC The polynucleotides and encoded polypeptides are potential markers for
 CC detecting, diagnosing, monitoring, characterising treating and
 CC potentially preventing breast cancer. The polynucleotides and encoded
 CC polypeptides are also useful for isolating compounds with cytostatic
 CC activity.
 XX
 SQ Sequence 807 BP; 90 A; 163 C; 125 G; 418 T; 11 other;
 Query Match 5.7%; Score 36.4; DB 22; Length 807;
 Best Local Similarity 44.0%; Pred. No. 1;
 Matches 133; Conservative 0; Mismatches 177; Indels 0; Gaps 0;
 QY 240 AGGAGATAGAACCTTGAACCCGACCAATAGCCGGTATGGAAGCGGATAGAACGCTCTAA 299
 DB 743 AGAAAGAAAAAGAGGGAAGGGAAGGGAAGGGAAGGGAAGGGAAGGGAAGGGAAGGGAAG 684
 QY 300 ATTTCACCATTCCTCCGACAGCGTTGGTTCACGACCTCACTAAACGCTTAAACGCGATAGAA 359
 DB 683 AATCAAGAGAGAAAG 624
 QY 360 GACAGAGATGGACCGCTAATAGAGAAACCTCGAGAGATAGAACCGCTCATCTACTAAAGTGA 419
 DB 623 GAGAAAAAGGGAAG 564
 QY 420 CAGATTCACAGTGTCTGTAGCTTAACGCCGACCAATAGCCCGACCAAGCGAGAGAGAGAGAG 479
 DB 563 AAGAAAAAGGGAAG 504
 QY 480 AAATATAGGACTCATTTGATCTAG 539
 DB 503 AAAAAAANANAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAA 444
 QY 540 CTAATAGTAGCAGTA 555
 DB 443 AAAAAAANAAAAA 428

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 24, 2002, 22:16:24 ; Search time 3529.69 Seconds
(without alignments)
2945.728 Million cell updates/sec

Title: US-10-068-080-11
Perfect score: 642
Sequence: 1 tactaatgtatcaaaataa.....tagttctaggattttatt 642

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- EST:*
- 1: em_estba:*
 - 2: em_esthum:*
 - 3: em_estin:*
 - 4: em_estnu:*
 - 5: em_estov:*
 - 6: em_estpl:*
 - 7: em_estro:*
 - 8: em_hic:*
 - 9: gb_est1:*
 - 10: gb_est2:*
 - 11: gb_hic:*
 - 12: gb_est3:*
 - 13: gb_est4:*
 - 14: gb_est5:*
 - 15: em_estfun:*
 - 16: em_estom:*
 - 17: gb_gss:*
 - 18: em_gss_hum:*
 - 19: em_gss_inv:*
 - 20: em_gss_pln:*
 - 21: em_gss_vrt:*
 - 22: em_gss_fun:*
 - 23: em_gss_mam:*
 - 24: em_gss_mus:*
 - 25: em_gss_other:*
 - 26: em_gss_pro:*
 - 27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	50	7.8	1621	BE868376	BE868376 601444337
c 2	48.4	7.5	997	CNS005TE	AL060767 Drosophill
3	46	7.2	1007	CNS06X9S	AL419462 T3 end of
c 4	44.6	6.9	555	AQ577769	AQ577769 nbxb0091L
5	44.2	6.9	589	CNS0301Z	AL253412 Tetraodon
6	44.2	6.9	789	AG061290	AG061290 Pan trogl

7	44	6.9	494	13	BM641696
8	44	6.9	1077	12	BG495797
9	43.2	6.7	972	17	CNS002JV
10	43	6.7	827	17	CNS02156
11	42.6	6.6	1689	12	BE872484
12	42.6	6.6	1715	12	BF243003
c 13	42.4	6.6	751	17	AG077983
14	42.2	6.6	679	17	AZ650047
15	42	6.5	712	17	AG056512
16	41.8	6.5	1101	17	CNS0162H
c 17	41.6	6.5	391	14	H43083
18	41.4	6.4	485	12	BF347872
c 19	41.4	6.4	878	9	AL536095
20	41.2	6.4	636	13	BJ335989
c 21	41.2	6.4	858	17	CNS012GG
c 22	40.8	6.4	954	17	CNS0050Y
23	40.8	6.4	1034	12	BF346237
c 24	40.4	6.3	902	17	CNS0060P
25	40.4	6.3	946	17	CNS007YY
26	40.2	6.3	894	17	CNS018BG
c 27	40.2	6.3	983	13	BM415036
c 28	40	6.2	1101	17	CNS0183G
29	39.8	6.2	386	17	BH863527
30	39.8	6.2	1892	12	BG121866
31	39.6	6.2	728	17	BH728612
32	39.4	6.1	494	13	BJ360881
33	39.4	6.1	654	13	BJ328915
34	39.2	6.1	562	17	PT013K11R
35	39.2	6.1	812	17	AG031898
36	39.2	6.1	904	17	AG056296
c 37	39.2	6.1	933	17	AZ166708
38	39.2	6.1	1019	13	BI648486
c 39	39.2	6.1	1101	17	CNS017RP
40	39	6.1	972	17	AG045001
c 41	39	6.1	1101	17	CNS0039G
42	38.8	6.0	366	9	AU262852
c 43	38.8	6.0	660	17	AQ873871
44	38.6	6.0	428	13	BJ391809
c 45	38.6	6.0	910	17	CNS0060N

ALIGNMENTS

RESULT 1
BE868376
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

601444337F1 NIH_MGC_65 Homo sapiens CDNA clone IMAGE:3848529 5',
mRNA sequence.
BE868376
BE868376.1 GI:10317152
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1621)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaphs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LIAW9564 row: i column: 10
High quality sequence stop: 33.
Location/Qualifiers
1. .1621

source

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3848529"
/clone_lib="NIH_MGC_65"
/tissue_type="adenocarcinoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.8 kb. Library constructed by Life Technologies."
BASE COUNT      862 a 446 c 260 g 53 t
ORIGIN
Query Match      7.8%; Score 50; DB 12; Length 1621;
Best Local Similarity 55.18; Pred. No. 0.00093;
Matches 98; Conservative 0; Mismatches 80; Indels 0; Gaps 0;
QY 383 AGAACTCGAGATAGAACCGTCTACTACTAGACAGATACAGTGTGCTAGCTAAC 442
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 324 AGAAACCAACAGACAGACGACCCACACAGAAACAAAGAACATAGAGACGAGAAACA 383
QY 443 GCCACATACCCACAGACAGACGAGAAAGGTGAAACACCAATAGACTCATTTGATCTGA 502
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 384 ACAGACAGAAACCAACAGACGACCAATTAACAAAAAATAACAACCTAATAACACACAAA 443
QY 503 GAGAGAGTACCGATATAGATCGTGAACCATCAACTAATAGTAAAGCAGTATTCCT 560
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 444 AACACCAACAGATATGAAGAAGACAGACCATCAATCAATAGCAACAGAAACAT 501

```

```

RESULT 2
CNS005TE/c      997 bp DNA linear GSS 03-JUN-1999
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence TET3 end of BAC #
BACR12K22 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION
AL060767
VERSION
AL060767.1 GI:4943573
KEYWORDS
GSS.
SOURCE
Drosophila melanogaster.
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 997)
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoeawa and
Aaron Mamoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
source
1..997
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACR12K22"
/clone_lib="RPCI-98"
/notes="end : TET3"
BASE COUNT      89 a 99 c 13 g 258 t 538 others

```

```

ORIGIN
Query Match      7.5%; Score 48.4; DB 17; Length 997;
Best Local Similarity 17.8%; Pred. No. 0.0026;
Matches 62; Conservative 132; Mismatches 154; Indels 0; Gaps 0;
QY 208 GAAGCGCTAATAAAGTAGATGTTCCCAAGAGGAGAGTAGAAGACCTGAAACCGACAA 267
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 908 SAGGGRGGRGRRRAAARAKDARARARARARARARARARARARARARARARARARRR 849
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 268 TAGCCGCTATGTAAGGATAGAAACGCTCTAAATTTCCACCATTCCGACCGTGTGCTCA 327
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 848 RRRARGAGGRGRRGRRRRARARRRAGARARRRRARRRRARRRRRRRRRRR 789
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 328 CGACCTCACTAAACCTTAACCGGATAGAGACAGATGGAACGCTAATAGAGAAA 387
      || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 788 RGARRRRGRRRRGRRRRRRARRARRARARARARARARARARARARARRRRR 729
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 388 CCTCGAGATAGAACGCTCATACTAAAGTCACAGATCACAGTGTCTAGTAAACCGCA 447
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 728 RRRRRRGAGARARRRRARRRRARRAGAAARRRRRRRRRRRRRRRRRRRRRR 669
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 448 CAATAGCCCCAAGACGAGAAAGGTGAAACCAAAATAGGACTCATTCATCTAGAGAG 507
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 668 RRRRRGRMRTRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 609
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 508 AAGTAGCGATATAGATCGTGAACGATCAACAACTAATAGTAAACGAGTA 555
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 608 RRRGRRARGARRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRA 561
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

RESULT 3
CNS06X9S
LOCUS
DEFINITION
T3 end of clone AX0AA039F08 of library AX0AA from strain CBS 7064
of Pichia farinosa, genomic survey sequence.
ACCESSION
AL419462
VERSION
AL419462.1 GI:12202640
KEYWORDS
GSS.
SOURCE
Pichia farinosa.
ORGANISM
Pichia farinosa
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Pichia.
REFERENCE
1 (bases 1 to 1007)
Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
Bislotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
Wincker,P. and Weissenbach,J.
Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies
FEBS Lett. 487 (1), 3-12 (2000)
PUBMED
11152876
REFERENCE
2 (bases 1 to 1007)
de Montigny,J., Spehner,C., Souciet,J., Tekala,F., Dujon,B.,
Wincker,P., Artiguenave,F. and Potier,S.
Genomic exploration of the hemiascomycetous yeasts: 15. Pichia
sorbitophila
FEBS Lett. 487 (1), 87-90 (2000)
PUBMED
20584725
REFERENCE
3 (bases 1 to 1007)
Genoscope.
Direct Submission
Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Creteil, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces

```

lactis var. *lactis*, *Kluyveromyces marxianus* var. *marxianus*, *Pichia angusta*, *Debaryomyces hansenii* var. *hansenii*, *Pichia sorbitophila*, *Candida tropicalis* and *varrowia lipolytica*. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

FEATURES

Location/Qualifiers
1..1007
/organism="Pichia farinosa"
/strain="CBS 7064"
/db_xref="taxon:4920"
/clone="AX0A039F08"
/clone_lib="AX0AA"
/note="end : T3"
BASE COUNT 533 a 86 c 126 g 161 t 101 others
ORIGIN

Query Match 7.2%; Score 46; DB 17; Length 1007;
Best Local Similarity 36.0%; Pred. No. 0.013;
Matches 121; Conservative 45; Mismatches 170; Indels 0; Gaps 0;
QY 214 TAATAAAAGTAGCTCCGCAAGAGGAGTAGAAGACCTGAAACCGACAATAGCGC 273
DB 164 WMAAANAATAAAGTTTAAAGAGAAAGAGAAATGAAATTAAMWMAAACA 223
QY 274 GTATGAAGGATAGAAACCTCTAAATTTCCACCATTCGGACAGCGTTGGTCAGCACCT 333
DB 224 MGTTCATGTCWGGGTGATGAMGTATWAGAWAAARAWARATTAAGGWTAAAGMA 283
QY 334 CACTAAACCTTAACCGCGATAGACAGACAGATGGAAGCTTAATAGAAAGCACTCA 393
DB 284 WAAWAAAWAAGAGGWTATTAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAA 343
QY 394 GAGATAGAACCGCTACTACTAAAGTACAGATCACAGTGTCTAGCTAAACCGCACAATAG 453
DB 344 AAHAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAW 403
QY 454 CCCAAGACGAGAAAGGTGAAACCAATAGGACTCTATTGATCTAGAGAGAGATAG 513
DB 404 ATAAAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAW 463
QY 514 CGTATAGATGCTGAACGATCAACTAATAGTAA 549
DB 464 WAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAW 499

RESULT 4

AQ577769/c 555 bp DNA linear GSS 02-JUN-1999
LOCUS nbx0091L07f CUGI Rice BAC Library Oryza sativa genomic clone
DEFINITION nbx0091L07f, DNA sequence.

ACCESSION AQ577769
VERSION AQ577769.1 GI:4978254

KEYWORDS GSS.
SOURCE Oryza sativa.

ORGANISM

Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.

REFERENCE

1 (bases 1 to 555)
Wing,R.A. and Dean,R.A.
A BAC End Sequencing Framework to Sequence the Rice Genome
Unpublished (1998)

JOURNAL

COMMENT Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: TAATAGACTCTACTATAGG
Class: BAC ends
High quality sequence stop: 101.

FEATURES

source

Location/Qualifiers

1..555
/organism="Oryza sativa"
/strain="Japonica"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/clone="nbx0091L07f"
/clone_lib="CUGI Rice BAC Library"
/tissue_type="Leaf"
/lab_host="E. coli DH10B"
/note="vector: pBelobAC11; Site 1: HindIII; Site 2: HindIII; Rice is one of two most popular grains in the world. Half of the world population especially those inhabiting highly populated areas of the humid tropics and subtropics, rely on rice as their primary source of carbohydrate. Monocotyledonous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 Mbp (Arumuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from Oryza sativa, Nipponbare variety. The library contains 36,864 clones with an average insert size of 128.5 Kb providing 10.9 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %. Two high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening." 4 others
BASE COUNT 89 a 120 c 86 g 256 t

Query Match 6.9%; Score 44.6; DB 17; Length 555;
Best Local Similarity 54.6%; Pred. No. 0.033;
Matches 89; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 383 AGAAACCTCGAGATAGAACCGCTCATCTAAAGTGACAGATCACAGTCTCCGTAC 442
DB 481 AGGAAAAAGAGAGATGGAACATGGATAAAAAACAACAGCTCAAAAAAGATAGAAGA 422

QY 443 GCGACATAGCCCAAGAGAGAGAGGTGAAAAACCAAAATAGGACTCATTGATCTCA 502
DB 421 GTAGGACAAATAATAGAACGACAAAGAAATAATAAAAAAATGTTAAAGATAGTGT 362

QY 503 GAGAGAAAGTAGCGTAATAAGAAATCGTGAACGATCAAACTAATA 545
DB 361 AGGAAAGCTAAGAAAAAGAAATATGAGGAGAGATATAAAA 319

RESULT 5

CNS0301Z 589 bp DNA linear GSS 17-MAY-2000
LOCUS Tetraodon nigroviridis genome survey sequence T7 end of clone
DEFINITION 043121 of library G from tetraodon nigroviridis, genomic survey sequence.

ACCESSION AL253412
VERSION AL253412.1 GI:7974424

KEYWORDS GSS: genome survey sequence.
SOURCE Tetraodon nigroviridis.

ORGANISM

Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.

REFERENCE

1 (bases 1 to 589)
Ruest-Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissenbach,J.

TITLE

Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence

JOURNAL

REFERENCE 2 (bases 1 to 589)
Ruest-Crolius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,

```

Sequencing: M13Rev
LIBRARY
Vector      : pKS145
R.Site 1   : SacI
R.Site 2   : SacI
Location/Qualifiers
1..789
/oranism="Pan troglodytes"

```

BASE COUNT	ORIGIN	Query Match	Score	DB	Length
552 a	57 c	136 g	19 t	25 others	789:

[illegible]

RESULT 7	BM641696	17000687308318	A Cam.ad cDNA1	Anopheles gambiae	cdNA clone	linear	EST 26-FEB-2002
LOCUS	BM641696	19600449660073	5'	mRNA	sequence.		
DEFINITION	ACCESSION	BM641696	1	GI:18941207			
	VERSION	EST.					
	KEYWORDS	AFRICAN MALARIA MOSQUITO.					
	SOURCE	ANOPHELES GAMBIAE					
	ORGANISM	EUKARYOTA; METAZOA; ARTHROPODA; HEXAPODA; INSECTA; PTERYGOTA; NEOPTERA; ENDOPTERYGOTA; DIPTERA; NEMATOCERA; CULICOIDEA; ANOPHELES.					
REFERENCE	AUTHORS	1 (bases 1 to 494)					
		Holt, R.A., Lin, J.-J., Murphy, S.D., Evans, C.A., Kraft, C.L., Charlab, R., Collins, F.H., Venter, J.C. and Hoffman, S.L.					
TITLE		Celera Anopheles gambiae EST project					
JOURNAL		Unpublished (2002)					
COMMENT		Contact: Holt R.A.					
		Celera Genomics					
		45 W. Gude Dr., Rockville, MD 20850, USA					
		Tel: 2404533151					
		Fax: 2404534580					


```

Location/Qualifiers
1. 1715
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4106347"
/clone_lib="NIH_MGC_55"
/tissue_type="from acute myelogenous leukemia"
/lab_host="DH10B (Tl phage-resistant)"
/note="Sf11: bone marrow; Vector: pONR-LIB (Clontech);
Site_1: Sf11 (ggcccatggccc); Site_2: Sf11 (ggccataggccc
); Double-stranded cDNA was prepared from cell line RNA.
5' and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CACGCCATTATGGCC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGCCGAGGCCGCGCATG-dT(30)BN-3'
(where B = A, C, or G and N = A, C, G, or T). Average
insert size 1.65 kb (range 0.9-4.0 kb). 14/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA)."
694 a 239 c 468 g 313 t 1 others

```

BASE COUNT 694 a 239 c 468 g 313 t 1 others
ORIGIN

Qy	338	AAAAGCCCTAAACGCGGATAGAAAGACAGACAGATGGGAACGCTAATAGAGAGAAACCTCGAGAGA	397
Db	1132	AGAAGACACACAGACAGAGAGAAGGCAAGAAGGAGCAGAAGGCGAAGCAAGAAGGAAGAGA	1191
Qy	398	TAGAACCCTCATACTAAAGTCACAGATCACACAGTGTCTAGCTTAACGCCGCAACAATAGCCCC	457
Db	1192	GGAAGAGAGTGAAGCAGAGAGAGAGAGGACCGGAGGAGCAGAGACAGCGCGAGAAAGGAGG	1251
Qy	458	AAGACGAGAAAGGTCGAAAAACAAAATAGGACTCATTTGATCTGAGAGAGAAAGTAGCGAT	517
Db	1252	AAGA- - -ATAGGAGAAAAAAGAAACGACGAGGACAGACAAAAACAAACAGGAAGAGAT	1307
Qy	518	AATAGAACTGCTGAACGATCAACCTAANTAGTA	548
Db	1308	AAAAGACAGAAAAAAGAGAAAGAAAGAAAGAGAA	1338

DEFINITION	Pan troglodytes male lymphoblast DNA, clone: PTB-072K12.F, genomic survey sequence.
LOCUS	Asu7983
ACCSSION	AG077983
VERSION	AG077983.1
KEYWORDS	GI:1629785
SOURCE	GSS.
ORGANISM	Pan troglodytes male lymphoblast DNA, clone_lib:PTB Chimpanzee Male BAC Library clone:PTB-072K12.F. Pan troglodytes
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.

TITLE	BAC end sequences of Library PTB
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 751)
AUTHORS	Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
TITLE	Direct Submission
JOURNAL	Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: chimpbes@gsc.riken.go.jp , URL: http://hgp.gsc.riken.go.jp/ , Tel: 81-45-503-9111, Fax: 81-45-503-9170)
COMMENT	Clones are derived from the chimpanzee BAC library PTB. This BAC end was generated during the R&D process and may have higher chance of c-one tracking errors.
	PRIMERS


```

REFERENCE
AUTHORS      1
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE        BAC end sequences of Library PTB
JOURNAL      Unpublished
REFERENCE    2 (bases 1 to 712)
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE        Direct Submission
JOURNAL      Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimpbes@sc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT      Clones are derived from the chimpanzee BAC library PTB This BAC end
was generated during the R&D process and may have higher chance of
clone tracking errors.
PRIMERS
Sequencing: M13Rev
LIBRARY
Vector       : pKS145
R.Site 1     : SacI
R.Site 2     : SacI
FEATURES
source       Location/Qualifiers
1..712
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone="PTB-042N03.R"
/sex="male"
/cell_type="lymphoblast"
/clone_lib="PTB Chimpanzee Male BAC Library"
BASE COUNT   403 a 66 c 218 g 15 t 10 others
ORIGIN

Query Match      6.5%; Score 42; DB 17; Length 712;
Best Local Similarity 45.9%; Pred. No. 0.2;
Matches 144; Conservative 0; Mismatches 170; Indels 0; Gaps 0;

Qy 236 AAAGAGGAGGTAGAAACCTCAAACCGACAAATAGCGGTATGGAAGGGATAGAAACGTC 295
      ||||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 255 AAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 314
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 296 CTAATTTCCACCATTCGACAGCGTTGGTCACGACCTCCTAAAGCCTAAACCGCGAT 355
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 315 AAAAAATCAAAAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 374
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 356 AGAAGACAGAGATGGAACCTAATAGAGAAACCTCGAGAGATAGAACCGTCATACTAAA 415
      ||| || | | | | | | | | | | | | | | | | | | | | | | | | |
Db 375 AGAGGAAAGAGAGAGAAAGAAACAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAG 434
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 416 GTGACAGATCAGTGTCTAGCTAACGCCGACAAATAGCCCAAGACGAGAAAGGTGAAA 475
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 435 GGGGAGAAAGAGAGAGAGAGAAAGAGGAGGAAAGAAAGAAAGAAAGAAAGAGGGGAG 494
      || | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 476 AACCAAAATAGACTATTGTACTTGAGAGAGAAAGTAGCGGATATAGAAATCGTGAACGAT 535
      || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 495 AAAAAAGAAAGAGAGAGAAAGAAAGAAAGAAAGAGAGAAAGAAAGAAAGAGAGAGAA 554
      || | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy 536 CAAACTAATAGTAA 549
      || | | | | | |
Db 555 CAAAGAACCAAGAAA 568

```

Search completed: December 25, 2002, 14:17:23
 Job time : 3537.69 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 24, 2002, 22:25:54 ; Search time 93.625 Seconds
(without alignments)
2102.926 Million cell updates/sec

Title: US-10-068-080-11
Perfect score: 642
Sequence: 1 tactaatgttatcaaaataa.....tagttctaggattttttatt 642

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A-COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B-COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A-COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B-COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PTUS-COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	327.8	51.1	344	US-09-222-938A-15	Sequence 15, Appl
2	100	15.6	582	US-09-222-938A-72	Sequence 72, Appl
3	67.2	10.5	7218	US-08-232-463-14	Sequence 14, Appl
4	38.2	6.0	5613	US-08-463-418-1	Sequence 1, Appl
5	33	5.1	597	US-08-332-766A-19	Sequence 19, Appl
6	33	5.1	1488	US-08-875-972-3	Sequence 3, Appl
7	33	5.1	2275	US-08-743-637B-2	Sequence 2, Appl
8	33	5.1	2275	US-08-526-840B-2	Sequence 2, Appl
9	33	5.1	2764	US-08-923-454A-9	Sequence 9, Appl
10	33	5.1	2764	US-08-832-867-2	Sequence 2, Appl
11	33	5.1	2765	US-08-888-077A-1	Sequence 1, Appl
12	33	5.1	2765	US-08-706-344C-1	Sequence 1, Appl
13	33	5.1	2765	US-08-706-344C-3	Sequence 3, Appl
14	33	5.1	2765	US-08-706-344C-27	Sequence 27, Appl
15	33	5.1	2765	US-08-706-344C-29	Sequence 29, Appl
16	33	5.1	2765	US-08-706-344C-31	Sequence 31, Appl
17	33	5.1	2791	US-08-967-101-1	Sequence 1, Appl
18	33	5.1	2791	US-08-967-101-133	Sequence 133, Appl
19	33	5.1	2791	US-08-592-541-1	Sequence 1, Appl
20	33	5.1	2791	US-08-592-541-133	Sequence 133, Appl
21	33	5.1	2791	US-09-124-698-1	Sequence 1, Appl
22	33	5.1	2791	US-09-124-698-133	Sequence 133, Appl
23	33	5.1	2791	US-09-127-480-1	Sequence 1, Appl
24	33	5.1	2791	US-09-127-480-133	Sequence 133, Appl
25	33	5.1	2791	US-08-496-841C-1	Sequence 1, Appl
26	33	5.1	2791	US-09-124-523-1	Sequence 1, Appl
27	33	5.1	2791	US-09-124-523-133	Sequence 133, Appl

C 28	33	5.1	2792	4	US-08-496-841C-133	Sequence 133, Appl
C 29	33	5.1	3086	3	US-08-888-077A-3	Sequence 3, Appl
C 30	33	5.1	3087	2	US-08-967-101-5	Sequence 5, Appl
C 31	33	5.1	3087	2	US-08-592-541-5	Sequence 5, Appl
C 32	33	5.1	3087	3	US-09-124-698-5	Sequence 5, Appl
C 33	33	5.1	3087	4	US-08-496-841C-5	Sequence 5, Appl
C 34	33	5.1	3087	4	US-08-496-841C-5	Sequence 5, Appl
C 35	33	5.1	3087	4	US-08-496-841C-5	Sequence 5, Appl
C 36	32.5	5.1	87350	3	US-08-781-891-79	Sequence 79, Appl
C 37	32.5	5.1	87543	4	US-09-791-211-3	Sequence 3, Appl
C 38	32.4	5.0	594	4	US-09-385-982-75	Sequence 75, Appl
C 39	32.4	5.0	3282	4	US-09-315-793-51	Sequence 51, Appl
C 40	32.4	5.0	4206	4	US-09-302-620B-81	Sequence 81, Appl
C 41	32	5.0	448	2	US-08-967-101-106	Sequence 106, Appl
C 42	32	5.0	448	2	US-08-592-541-106	Sequence 106, Appl
C 43	32	5.0	448	3	US-09-124-698-106	Sequence 106, Appl
C 44	32	5.0	448	4	US-09-127-480-106	Sequence 106, Appl
C 45	32	5.0	448	4	US-08-496-841C-106	Sequence 106, Appl

ALIGNMENTS

RESULT 1
US-09-222-938A-15
; Sequence 15. Application US/09222938A
; Patent No. 6437108
; GENERAL INFORMATION:
; APPLICANT: Youngman, Philip
; APPLICANT: Fritz, Christian
; APPLICANT: Murphy, Christopher
; APPLICANT: Guzman, Luz-Maria
; TITLE OF INVENTION: ESSENTIAL BACTERIAL GENES AND THEIR USE
; FILE REFERENCE: 07334/060001
; CURRENT APPLICATION NUMBER: US/09/222,938A
; CURRENT FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 15
; LENGTH: 344
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-09-222-938A-15

Query Match	51.1%	Score	327.8;	DB 4;	Length	344;			
Best Local Similarity	97.9%;	Pred. No.	3.1e-95;						
Matches	332;	Conservative	0;	Mismatches	7;	Indels	0;	Gaps	0;
Qy	113	TGTGACCTTGGTGTGGTGTGGAAGCGTAAATCCATTTCTTCGACCATACCGTTGGAAAC	172						
Db	6	TGTGACCTTGGTGTGGTGTGGAAGCGTAAATCCATTTCTTCGACCATACCGTTGGAAAC	65						
Qy	173	ACTAATCTGAAAAAGTTTCCCTTGGGATCGTTGCGACGAAAGGCTAATAAAAACTAGATGTTTC	232						
Db	66	ACTAATCTGAAAAAGTTTCCCTTGGGATCGTTGCGACGAAAGGCTAATAAAAACTAGATGTTTC	125						
Qy	233	CGCAAGAGGAGAGTAGAAACCTGAAACCGACCAATAGCCGGTATGGAAGGGATAGAAAC	292						
Db	126	CGCAAGAGGAGAGTAGAAACCTGAAACCGACCAATAGCCGGTATGGAAGGGATAGAAAC	185						
Qy	293	GTCTTAATTTCCACCAATTCGACAGCGTGTGGTCACGACCTCACTAAAAGCCTAAACCGG	352						
Db	186	GTCTTAATTTCCACCAATTCGACAGCGTGTGGTCACGACCTCACTAAAAGCCTAAACCGG	245						
Qy	353	GATGAAGACAGAGATGGAACGCTAATAGAAGAAACCTTCGAGAGATGAACCGGTCATACT	412						
Db	246	GATGAAGACAGAGATGGAACGCTAATAGAAGAAACCTTCGAGAGTATAGAACCCTACTACT	305						
Qy	413	AAAGTGACAGATCACAGTGTGCTAGCTAAACGCCGACAAAT	451						
Db	306	AAAGTGACAGATCACAGTGTGCTAGCTAGCCGACGACAAAT	344						

RESULT 2

US-09-222-938A-72
; Sequence 72, Application US/09222938A
; Patent No. 6437108
; GENERAL INFORMATION:
; APPLICANT: Youngman, Philip
; APPLICANT: Fritz, Christian
; APPLICANT: Murphy, Christopher
; APPLICANT: Guzman, Luz-Maria
; TITLE OF INVENTION: ESSENTIAL BACTERIAL GENES AND THEIR USE
; FILE REFERENCE: 07334/060001
; CURRENT APPLICATION NUMBER: US/09/222.938A
; CURRENT FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 72
; LENGTH: 582
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-09-222-938A-72

Query Match 15.6%; Score 100; DB 4; Length 582;
Best Local Similarity 52.9%; Pred. No. 1.9e-22;
Matches 239; Conservative 0; Mismatches 210; Indels 3; Gaps 1;

Qy 1 TACTAATGCTTATCAAAATAATTAGCATCGGATACACACCAAGCTAAGGTAGACACAGAG 60
Db 1 TACTAATTAACCAATAACTAATAAACCGGATGAACCTATCGTCTGTAAGGTAGACCAAGT 60

Qy 61 ACCTAACCTGTTCTAATAAGAAATTTAGTTAGATCGCTGCTTACCAAGACCAATTGTGACCT 120
Db 61 TAACACCGCTGCGAAGCTTTCTTAACATAAGCCCTCGTGCCTTCGCCCTTGAATCCG 120

Qy 121 TGCTGGTGTGGAAGGCTAAATTCATTTTCGACATACACCGTTGGAAACACTAAGT 180
Db 121 CGATGGTTACGTAAGGCTATGAACCAATTTTCGACCAAGCCAGCAGTATCGGCCCTCTA 180

Qy 181 AAAAGTTTCCTCGGATCGTTGCGACGAAGGCTAATAAAGATAGATGTTCCGCAAGA 240
Db 181 TAAACATTTTCCTGACCGTTGACGTAACGGAAGAGTACGTACA---ACTATAAGTG 237

Qy 241 GGAGATGAGAAACCTGAAACCGCAATAGCGGTATGGAAGGATAGAAACGTCCTTAA 300
Db 238 GCGAAGAAGCTGCTCAGAAACGCAAAATCCGCTGCAAAAGGATGAAGCGGTTTAA 297

Qy 301 TTTCCACATTCGACACGCTGTCAGGACCTCACTAAAGCCCTAAACGCGGATAGAG 360
Db 298 TTTCCGCAATTCGCGACCGCTGCTAGTCTCCGCAAAACGATAAAATGCGTGGGCAAT 357

Qy 361 ACAGAGATGGAACGCTAATAAGAAACCTCGAGAGATAGAACGCTCATACTAAGTGAC 420
Db 358 AAATAGTCTACCAACGCCATAAGAGTAGAAAATATGAAGTATTAACAAAGAGAG 417

Qy 421 AGATCAGCTGCTGCTAGTACGCGGCAATA 452
Db 418 AGTAGCTACAATTGTCCTAGATATGACAATA 449

RESULT 3

US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA

COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZ9pt-Fls
US-08-232-463-14

Query Match 10.5%; Score 67.2; DB 1; Length 7218;
Best Local Similarity 5.3%; Pred. No. 1.8e-11;
Matches 21; Conservative 226; Mismatches 149; Indels 0; Gaps 0;

Qy 158 CATACCGTTGGAACACTAAGTTCCTTGGATCTTCGCGAGCAAGCGTAAT 217
Db 1476 CCTATCTATCAAGTAGTTAAAGAGATAGAAGAAATTTGGTACRRRRRRRRRRRR 1417

Qy 218 AAAAAGTAGATGTTCCGCAAGAGAGAGTAGAAACCTGAAACCGACATAAGCCGTAT 277
Db 1416 RRR 1357

Qy 278 GGAAGGATAGAACGCTCTAAATTTCCACCATTCGCGACAGCGTTGGTCAAGCCTCACT 337
Db 1356 RRR 1297

Qy 338 AAAAGCCTAAACCGGATAGAACACAGAGATGCAACGCTAATAGAAACCTCGAGAGA 397
Db 1296 RRR 1237

Qy 398 TAGAACCGCTACTATAAGTGACAGATCACAGTGTGCTAGCTAACGCGCAATAGCC 457
Db 1236 RRR 1177

Qy 458 AAGACGAGAAAGTGAGAAACCAAAATAGGACTCATTTGATCTGAGAGAGAGTAGCGAT 517
Db 1176 RRR 1117

Qy 518 AATAGATCTGGAACGATCAAACTAATAGTAACGAG 553
Db 1116 RRR 1081

RESULT 4

US-08-463-418-1
; Sequence 1, Application US/08463418
; Patent No. 5908971
; GENERAL INFORMATION:
; APPLICANT: Van Der Straeten, Dominique et al.

;/ TITLE OF INVENTION: CRUCIFER ACC SYNTHASE AND USES THEREOF
;/ NUMBER OF SEQUENCES: 10
;/ CORRESPONDENCE ADDRESS:
;/ ADDRESSEE: Fish & Richardson P.C.
;/ STREET: 225 Franklin Street
;/ CITY: Boston
;/ STATE: MA
;/ COUNTRY: USA
;/ ZIP: 02110-2804
;/ COMPUTER READABLE FORM:
;/ MEDIUM TYPE: Floppy disk
;/ COMPUTER: IBM PC compatible
;/ OPERATING SYSTEM: PC-DOS/MS-DOS
;/ SOFTWARE: Patent In Release #1.0, Version #1.30
;/ CURRENT APPLICATION DATA:
;/ APPLICATION NUMBER: US/08/463,418
;/ FILING DATE: 05-JUN-1995
;/ CLASSIFICATION: 800
;/ PRIOR APPLICATION DATA:
;/ APPLICATION NUMBER: US 07/962,481
;/ FILING DATE: 15-OCT-1992
;/ ATTORNEY/AGENT INFORMATION:
;/ NAME: Clark, Paul T.
;/ REGISTRATION NUMBER: 30,162
;/ REFERENCE/DOCKET NUMBER: 00786/161002
;/ TELECOMMUNICATION INFORMATION:
;/ TELEPHONE: 617/542-5070
;/ TELEFAX: 617/542-8906
;/ TELEX: 200154
;/ INFORMATION FOR SEQ ID NO: 1:
;/ SEQUENCE CHARACTERISTICS:
;/ LENGTH: 5613 base pairs
;/ TYPE: nucleic acid
;/ STRANDEDNESS: double
;/ TOPOLOGY: linear
;/ MOLECULE TYPE: cDNA
;/ US-08-463-418-1

Query Match 6.0%; Score 38.2; DB 2; Length 5613;
Best Local Similarity 48.8%; Pred. No. 0.028;
Matches 103; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

QY 375 CTAATAGAGAAACCTCGAGAGATAGAACCGTCTACTACTAAAGTGACAGATCAGCTGTCG 434
Db 434 CTAACAAAAAAGAAATCAAGATAAATAAATACTCTCACTAGTCTCACTAATAATTTTC 493
QY 435 TAGCTAACCCGCNATACCCCAACAGACGAGAAAGGTGAAACCAAAATAGGACTCAT 494
Db 494 ATTACAATTTCATATATGATCTACTACTAACAATTGTTGGTTATACAAACAAAAATTTAT 553
QY 495 GATACCTGAGAGAGAACTAGCAGATAGATAGTCTGAAACGATCAAACTAATAGTAAGCACT 554
Db 554 TTTCGTAAGACGGAATTTTGAATCAAAATCTCAGACTCAATGATATTCAGTAGT 613
QY 555 ATTCCTGTTATATCGAGCATAGTTTTTTTATTT 585
Db 614 AGTTTATCAATGACTAGATTAGATATTTCTT 644

RESULT 5
US-08-332-766A-19
; Sequence 19, Application US/08332766A
; Patent No. 5843647
; GENERAL INFORMATION:
; APPLICANT: JEFFREYS, Alec J.
; APPLICANT: ARMOUR, John
; TITLE OF INVENTION: SIMPLE TANDEM REPEATS
; NUMBER OF SEQUENCES: 125
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.
; STREET: 1100 New York Avenue, N.W.
; CITY: Washington
; STATE: D. C.

;/ COUNTRY: U.S.A.
;/ ZIP: 20005-3918
;/ COMPUTER READABLE FORM:
;/ MEDIUM TYPE: Floppy disk
;/ COMPUTER: IBM PC compatible
;/ OPERATING SYSTEM: PC-DOS/MS-DOS
;/ SOFTWARE: Patent In Release #1.0, Version #1.25
;/ CURRENT APPLICATION DATA:
;/ APPLICATION NUMBER: US/08/332,766A
;/ FILING DATE: 01-NOV-1994
;/ CLASSIFICATION: 435
;/ PRIOR APPLICATION DATA:
;/ APPLICATION NUMBER: GB 9326052.9
;/ FILING DATE: 21-DEC-1993
;/ ATTORNEY/AGENT INFORMATION:
;/ NAME: BIRD, Donald J.
;/ REGISTRATION NUMBER: 25,323
;/ REFERENCE/DOCKET NUMBER: 217211/M94/0434/GB
;/ TELECOMMUNICATION INFORMATION:
;/ TELEPHONE: (202) 861-3000
;/ TELEFAX: (202) 822-0944
;/ TELEX: 6714627 CUSH
;/ INFORMATION FOR SEQ ID NO: 19:
;/ SEQUENCE CHARACTERISTICS:
;/ LENGTH: 597 base pairs
;/ TYPE: nucleic acid
;/ STRANDEDNESS: single
;/ TOPOLOGY: linear
;/ MOLECULE TYPE: DNA (genomic)
;/ US-08-332-766A-19

Query Match 5.1%; Score 33; DB 2; Length 597;
Best Local Similarity 45.3%; Pred. No. 0.46;
Matches 120; Conservative 0; Mismatches 145; Indels 0; Gaps 0;

QY 336 CTTAAAGCCTAAACCGGATAGAGACAGAGATGGAACCGCTTAATAGAGAAACCTCGAGA 395
Db 232 CTTAAAGCTGAAAGAAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGA 291
QY 396 GATAGAACCGTCTACTACTAAAGTGACAGATCAGAGTCTGCTAGCTAACGCCGACATAGCC 455
Db 292 GAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 351
QY 456 CCAGACGACGAGAAAGGTGAAAAACCAAAATAGGACTCATTGATCTAGAGAGAGAGTAGCG 515
Db 352 AATAAG 411
QY 516 APTATAGATCTGTAACCATCAAACTAATAGTAAGCAGATTTCTCTGTTATATCGAGCATA 575
Db 412 AGAAAG 471
QY 576 GTTTTATTTTTCACCTTTTAAACCCAG 600
Db 472 GTTGGCTTTTTTTTTTTTCTGTGTCAG 496

RESULT 6
US-08-875-972-3/G
; Sequence 3, Application US/08875972
; Patent No. 5985564
; GENERAL INFORMATION:
; APPLICANT: Huntington Potter and Jinhue Li
; TITLE OF INVENTION: ASSAY FOR IDENTIFYING GENES CAUSING
; TITLE OF INVENTION: CHROMOSOME NON-DISJUNCTION
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HAMILTON, BROOK, SMITH & REYNOLDS, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173-4799
; COMPUTE READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/75,972
FILING DATE: 08-AUG-97
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/002,448
FILING DATE: 16-AUG-1995
ATTORNEY/AGENT INFORMATION:
NAME: Granahan Esq., Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: H095-03PA
TELEPHONE: (781) 861-6240
TELEFAX: (781) 861-9540
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1488 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 2..1222
US-08-875-972-3

Query Match 5.1%; Score 33; DB 2; Length 1488;
Best Local Similarity 55.6%; Pred. No. 0.71; Mismatches 0; Indels 65; Gaps 1;
Matches 85; Conservative 0;
QY 111 ATTGTGACCTGCTGTTGTGGAAGCGTAAATCCATCTTTTCGACCATACCGTTGGAA 170
DB 1329 ATCTGACTTTGTTAGATGGACACAGGAAATACCTTTGTCTCCCGCAGATTGGTT 1270
QY 171 ACCTAACTGAAAGCTTCCCTTGGGATCGTTGCGAGAA---GGCTAATAAACTAGA 227
DB 1269 ATAGTCAAGAGAAATCCATGGGATCTTAACCCCAATATGCTAGATATAAAATTGA 1210
QY 228 TGTCCGCAAGAGGAGACTAGAAACCTGAAAA 260
DB 1209 TGCAATGCTAATTTGGTCCATAAAAGCGTGTACA 1177

RESULT 7
US-08-743-637B-2
Sequence 2, Application US/08743637B
Patent No. 5994066
GENERAL INFORMATION:
APPLICANT: BERGERON, Michel G.
APPLICANT: PICARD, Francois J.
APPLICANT: OUELLETTE, Marc
APPLICANT: ROY, Paul H.
TITLE OF INVENTION: SPECIES-SPECIFIC AND UNIVERSAL DNA
TITLE OF INVENTION: PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND
TITLE OF INVENTION: IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED
TITLE OF INVENTION: ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ...
NUMBER OF SEQUENCES: 273
CORRESPONDENCE ADDRESS:
ADDRESSEE: QUARLES & BRADY
STREET: 411 EAST WISCONSIN AVENUE
CITY: MILWAUKEE
STATE: WISCONSIN
COUNTRY: USA
ZIP: 53202-4497
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/743,637B
FILING DATE: 04-NOV-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/526,840
FILING DATE: 11-SEP-1995
ATTORNEY/AGENT INFORMATION:
NAME: BAKER, Jean C.
REGISTRATION NUMBER: 35,433
REFERENCE/DOCKET NUMBER: 850586.90012
TELEPHONE: (414) 277-5000
TELEFAX: (414) 277-5591
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2275 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecalis
US-08-743-637B-2
Query Match 5.1%; Score 33; DB 2; Length 2275;
Best Local Similarity 49.7%; Pred. No. 0.86;
Matches 84; Conservative 0; Mismatches 85; Indels 0; Gaps 0;
QY 14 AAAATAATTAGGATCGATAGACGACCCCAAGCTAAGGTAGACAGACCTAACCTGTTC 73
DB 1612 AATGGATTTTATGATGACAGACATGCGACCTATATTGAAAAAATTATCAACCGTTCC 1671
QY 74 ATAAGAAAGTTAGTTAGATCGCTGCTACCAAGACCATTTGTGACCTTGCTGGTTGGA 133
DB 1672 AAAATAAATATATTTTCAGATGCTATTACACGTGTAGACGGACCAATCAGAAATTAG 1731
QY 134 AGCGCTAAATCATCTTTCCACCATACCGTTGGAAACACTAACTGAA 182
DB 1732 GTGCGCAAGACGGTTTATTCGCCAATCCGTGAATTACAGGACGCA 1780

RESULT 8
US-08-526-840B-2
Sequence 2, Application US/08526840B
Patent No. 6001564
GENERAL INFORMATION:
APPLICANT: BERGERON, Michel G.
APPLICANT: OUELLETTE, Marc
APPLICANT: ROY, Paul H.
TITLE OF INVENTION: SPECIFIC AND UNIVERSAL PROBES AND
TITLE OF INVENTION: AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND IDENTIFY
TITLE OF INVENTION: COMMON BACTERIAL PATHOGENS AND ANTIBIOTIC RESISTANCE GEN
NUMBER OF SEQUENCES: 177
CORRESPONDENCE ADDRESS:
ADDRESSEE: QUARLES & BRADY
STREET: 411 East Wisconsin Avenue
CITY: Milwaukee
STATE: Wisconsin
COUNTRY: USA
ZIP: 53202-4497
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/526,840B
FILING DATE: 11-SEP-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/304,732

FILING DATE: 12-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: BAKER, Jean C.
REGISTRATION NUMBER: 35,433
REFERENCE/DOCKET NUMBER: 850586.90012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (414) 277-5000
TELEFAX: (414) 277-5591
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2275 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecalis

US-08-526-840B-2
Query Match 5.1%; Score 33; DB 3; Length 2275;
Best Local Similarity 49.7%; Pred. No. 0.86;
Matches 84; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

Qy 14 AAATAATTAGTCGGATAGACGACCAAGCTAAGTAGACAGACCTAACTGTTTC 73
|| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 1612 AATGGAATTTGATGACCAAGAAGCATGCAGCCTATTATGAAAAAATTATCAACGGTTTCC 1671
Qy 74 ATAGAAAGTTTACTAGATCGCTCTGACCAAGACCACTTCTGACCTTGCTGGTGTGGA 133
|| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 1672 AAATAAATATTTTCAGATGCTATTACACGCTGACGACCAATCAAGAAATAG 1731
Qy 134 AGGCGTAAATCCATTTCTTGACCATACCTGTTGGAACACATAACTGAA 182
|| || || || || || || || || || || || || || || || || || || || ||
Db 1732 GTGCCCAAGACGGTTTATTGACCAATCCGCTGAATTACAGGAACGCA 1780

RESULT 9
US-08-923-454A-9/c
Sequence 9, Application US/08923454A
Patent No. 6004794
GENERAL INFORMATION:
APPLICANT: Creasy, Caretha
APPLICANT: Livi, George
APPLICANT: Kattan, Eric
APPLICANT: Clindenbeard, Helen
APPLICANT: Browne, Michael
APPLICANT: Southan, Christopher
TITLE OF INVENTION: HUMAN SERINE PROTEASE
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/923,454A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/025436
FILING DATE: 06-SEPT-1996
ATTORNEY/AGENT INFORMATION:
NAME: Baumeister, Kirk
REGISTRATION NUMBER: 33,833
REFERENCE/DOCKET NUMBER: P50547
TELECOMMUNICATION INFORMATION:

TELEPHONE: 610-270-5096
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 2764 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE:
ORIGINAL SOURCE:
US-08-923-454A-9

Query Match 5.1%; Score 33; DB 3; Length 2764;
Best Local Similarity 55.6%; Pred. No. 0.94;
Matches 85; Conservative 0; Mismatches 65; Indels 3; Gaps 1;

Qy 111 ATTGTGACCTTGCCTGGTGTGGAGCGGTAAATCCATTCTTTGACCATACCTGTGAA 170
|| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 1756 ATCTTGACTTTGTAGATGTGGACAGAGAAATACCTTTGTCTCCCGAGATTGGTT 1697
Qy 171 ACATTAACCTGAAAGTTTCTTGGGATCGTTGCGACGAA---GGCTAATAAAAGTAGA 227
|| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 1696 ATACTCAAGAAGAAACATCCATGGGATTCTAACCGCAATATGCTAGATATAAATGA 1637
Qy 228 TGTTCGCCAAGAGGAGTAGAAGACCTGAAA 260
|| || || || || || || || || || || || || || || || || || || || ||
Db 1636 TGGATGCTAATTTGGTCCATATAAAAGGCTGTACA 1604

RESULT 10
US-08-832-867-2/c
Sequence 2, Application US/08832867C
Patent No. 6376239
GENERAL INFORMATION:

APPLICANT: BAUMEISTER, Ralf
TITLE OF INVENTION: DNA MOLECULES COMPRISING A PROMOTER CAPABLE OF
TITLE OF INVENTION: CONFERRING EXPRESSION OF A HETEROLOGOUS DNA SEQUENCE IN
TITLE OF INVENTION: ALL NEURAL CELLS AT ALL STAGES OF DEVELOPMENT IN C.

FILE REFERENCE: 674503-2004
CURRENT APPLICATION NUMBER: US/08/832,867C
CURRENT FILING DATE: 1997-04-04
NUMBER OF SEQ ID NOS: 8
SOFTWARE: PatentIn ver. 2.0
SEQ ID NO 2
LENGTH: 2764
TYPE: DNA
ORGANISM: Caenorhabditis elegans
US-08-832-867-2

Query Match 5.1%; Score 33; DB 4; Length 2764;
Best Local Similarity 55.6%; Pred. No. 0.94;
Matches 85; Conservative 0; Mismatches 65; Indels 3; Gaps 1;

Qy 111 ATTGTGACCTTGCCTGGTGTGGAGCGGTAAATCCATTCTTTGACCATACCTGTGAA 170
|| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 1756 ATCTTGACTTTGTAGATGTGGACAGAGAAATACCTTTGTCTCCCGAGATTGGTT 1697
Qy 171 ACATTAACCTGAAAGTTTCTTGGGATCGTTGCGACGAA---GGCTAATAAAAGTAGA 227
|| || |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 1696 ATAGTCAAGAGAGAAACATCCATGGGATTCTAACCGCAATATGCTAGATATAAATGA 1637
Qy 228 TGTTCGCCAAGAGGAGTAGAAGACCTGAAA 260
|| || || || || || || || || || || || || || || || || || || || ||
Db 1636 TGGATGCTAATTTGGTCCATATAAAAGGCTGTACA 1604

RESULT 11
US-08-888-077A-1/c

STATE: DC
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/706,344C
FILING DATE: 30-AUG-1996
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/003,054
FILING DATE: 31-AUG-1995
ATTORNEY/AGENT INFORMATION:
NAME: KIM, JUDITH U.
REGISTRATION NUMBER: 40,679
REFERENCE/DOCKET NUMBER: 0609.4180001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2765 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 249..1649
US-08-706-344C-3

Query Match 5.1%; Score 33; DB 4; Length 2765;
Best Local Similarity 55.6%; Pred. No. 0.94;
Matches 85; Conservative 0; Mismatches 65; Indels 3; Gaps 1;

QY 111 ATTGTGACCTGCTGCTGGTGTGGAAGCGTAAATCCATCTTTCGACCATACCGTTGGAA 170
DB 1756 ATCTTGACCTTTGTTAGATGTGCACACAGGAAATCACCTTTGCTCTCCCGAGATTGGTT 1697

QY 171 ACACCTAACTGAAAAAGTTTCTTGGGATCGTTGGCAGCAA---GGCTAATAAAAGTAGA 227
DB 1696 ATAGTCAAGAGAAACATCCATGGATTCTAACCGCAATATGCTAGATATAAAATTGA 1637

QY 228 TGTCCGCAAGAGAGAGTAGAACCTGAAAA 260
DB 1636 TGGATGCTAATTGGTCCATAAAGGCTGTACA 1604

RESULT 14
US-08-706-344C-27/c
Sequence 27, Application US/08706344C
Patent No. 6248555
GENERAL INFORMATION:
APPLICANT: TANZI, RUDOLPH
APPLICANT: WASCO, WILMA
TITLE OF INVENTION: Genetic Alterations Related To Familial
TITLE OF INVENTION: Alzheimer's Disease
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 NEW YORK AVENUE, SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/706,344C
FILING DATE: 30-AUG-1996
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/003,054
FILING DATE: 31-AUG-1995
ATTORNEY/AGENT INFORMATION:
NAME: KIM, JUDITH U.
REGISTRATION NUMBER: 40,679
REFERENCE/DOCKET NUMBER: 0609.4180001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 2765 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 249..1649
US-08-706-344C-27

Query Match 5.1%; Score 33; DB 4; Length 2765;
Best Local Similarity 55.6%; Pred. No. 0.94;
Matches 85; Conservative 0; Mismatches 65; Indels 3; Gaps 1;

QY 111 ATTGTGACCTGCTGCTGGTGTGGAAGCGTAAATCCATCTTTCGACCATACCGTTGGAA 170
DB 1756 ATCTTGACCTTTGTTAGATGTGCACACAGGAAATCACCTTTGCTCTCCCGAGATTGGTT 1697

QY 171 ACACCTAACTGAAAAAGTTTCTTGGGATCGTTGGCAGCAA---GGCTAATAAAAGTAGA 227
DB 1696 ATAGTCAAGAGAAACATCCATGGATTCTAACCGCAATATGCTAGATATAAAATTGA 1637

QY 228 TGTCCGCAAGAGAGAGTAGAACCTGAAAA 260
DB 1636 TGGATGCTAATTGGTCCATAAAGGCTGTACA 1604

RESULT 15
US-08-706-344C-29/c
Sequence 29, Application US/08706344C
Patent No. 6248555
GENERAL INFORMATION:
APPLICANT: TANZI, RUDOLPH
APPLICANT: WASCO, WILMA
TITLE OF INVENTION: Genetic Alterations Related To Familial
TITLE OF INVENTION: Alzheimer's Disease
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 NEW YORK AVENUE, SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/706,344C
FILING DATE: 30-AUG-1996
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/003,054
FILING DATE: 31-AUG-1995
ATTORNEY/AGENT INFORMATION:

Search completed: December 25, 2002, 14:26:36
Job time : 110.625 secs

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 25, 2002, 11:18:11 : Search time 189.61 Seconds
(without alignments)
1375.466 Million cell updates/sec

Title: US-10-068-080-11
Perfect score: 642
Sequence: 1 tactaatgttatcaaaataa.....tagttctaggattttttatt 642

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 363474 seqs, 203117208 residues
Total number of hits satisfying chosen parameters: 726948

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : Published_Applications_NA.*
- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
 - 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
 - 3: /cgn2_6/ptodata/2/pubpna/US05_NEW_PUB.seq.*
 - 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
 - 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
 - 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
 - 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
 - 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
 - 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
 - 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
 - 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
 - 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
 - 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
 - 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	642	100.0	642	12	US-10-068-080-11
2	100	15.6	582	12	US-10-068-080-12
C 3	35.6	5.5	640681	10	US-09-790-988-1
C 4	34.2	5.3	180216	10	US-09-835-232-6
C 5	33.6	5.2	1058	9	US-09-938-842A-2917
6	33.4	5.2	236	10	US-09-864-761-33009
7	33.4	5.2	29729	10	US-09-864-761-16513
8	33.4	5.2	1155	10	US-09-815-242-6498
9	33	5.1	2275	10	US-09-452-599-2
10	33	5.1	2765	10	US-09-785-474-1
C 11	33	5.1	2765	10	US-09-785-474-3
C 12	33	5.1	2765	10	US-09-785-474-27
C 13	33	5.1	2765	10	US-09-785-474-29
C 14	33	5.1	2765	10	US-09-785-474-31
C 15	33	5.1	2765	10	US-09-785-474-31
C 16	32.6	5.1	691	10	US-09-878-574-4162
C 17	32.6	5.1	1125	10	US-09-974-300-4278
C 18	32.6	5.1	2744	10	US-09-070-927A-548
C 19	32.6	5.1	1503841	9	US-09-946-807-1

C 20	32.6	5.1	1503841	10	US-09-795-668-1	Sequence 1, Appli
C 21	32.6	5.1	1503841	10	US-09-795-668-1	Sequence 1, Appli
C 22	32.4	5.0	921	10	US-09-974-300-4778	Sequence 4778, Ap
C 23	32.4	5.0	1560	9	US-09-764-868-1432	Sequence 1432, Ap
C 24	32.4	5.0	4206	10	US-09-911-781-2	Sequence 2, Appli
C 25	32.4	5.0	7258	10	US-09-790-988-3	Sequence 3, Appli
C 26	32.2	5.0	396	10	US-09-878-574-1207	Sequence 1207, Ap
C 27	32.2	5.0	615	9	US-09-950-933A-28	Sequence 28, Appl
C 28	32.2	5.0	684973	10	US-09-263-959-1	Sequence 1, Appli
C 29	32	5.0	541	10	US-09-864-761-14853	Sequence 14853, A
C 30	32	5.0	2000	9	US-09-938-842A-4002	Sequence 4002, Ap
C 31	31.8	5.0	487	9	US-10-046-935-701	Sequence 701, App
C 32	31.8	5.0	487	9	US-09-878-178-701	Sequence 701, App
C 33	31.8	5.0	2000	9	US-09-938-842A-4834	Sequence 4834, Ap
C 34	31.6	4.9	31885	10	US-09-764-860-775	Sequence 775, App
C 35	31.6	4.9	31885	10	US-09-764-877-2530	Sequence 2530, Ap
C 36	31.6	4.9	31885	10	US-09-764-877-2541	Sequence 2541, Ap
C 37	31.4	4.9	493	10	US-09-895-035-3	Sequence 3, Appli
C 38	31.2	4.9	202	10	US-09-764-869-159	Sequence 159, App
C 39	31.2	4.9	464	10	US-09-783-590-9045	Sequence 9045, Ap
C 40	31.2	4.9	1968	9	US-09-991-496-130	Sequence 130, App
C 41	31.2	4.9	2000	9	US-09-938-842A-4256	Sequence 4256, Ap
C 42	31	4.8	385	10	US-09-920-300A-472	Sequence 472, App
C 43	31	4.8	385	12	US-10-033-528-472	Sequence 472, App
C 44	31	4.8	392	10	US-09-867-701-1908	Sequence 1908, Ap
C 45	31	4.8	440	10	US-09-998-598-1083	Sequence 1083, Ap

ALIGNMENTS

RESULT 1
US-10-068-080-11
; Sequence 11, Application US/10068080
; Patent No. US20020115591A1
; GENERAL INFORMATION:
; APPLICANT: Fritz, Christian
; APPLICANT: Youngman, Philip
; TITLE OF INVENTION: USE OF YNES, ESSENTIAL BACTERIAL GENES AND POLYPEPTIDES
; FILE REFERENCE: 07334/090001
; CURRENT APPLICATION NUMBER: US/10/068,080
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/070,116
; PRIOR FILING DATE: 1997-12-31
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 11
; LENGTH: 642
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-10-068-080-11

Query Match	100.0%	Score 642;	DB 12;	Length 642;
Best Local Similarity	100.0%;	Pred. No. 3.8e-181;		
Matches 642;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY 1	TACTAATGTTATCAAAATATTAGGATCGGATAGACGACCCAGCTAAGGTAGACAGAG 60			
Db 1	TACTAATGTTATCAAAATATTAGGATCGGATAGACGACCCAGCTAAGGTAGACAGAG 60			
QY 61	ACCTAACCTGTTTCATAGAAGTTTATGTTAGATCGCTCGTACCAAGACCATTTGTGACCT 120			
Db 61	ACCTAACCTGTTTCATAGAAGTTTATGTTAGATCGCTCGTACCAAGACCATTTGTGACCT 120			
QY 121	TGCTGTTGTTGGAAGCGTAAATCCATTTCTTCGACCATACCGTTTGGAAACACTAACGTG 180			
Db 121	TGCTGTTGTTGGAAGCGTAAATCCATTTCTTCGACCATACCGTTTGGAAACACTAACGTG 180			
QY 181	AAAAGTTTCCCTTGGGATCGTTCGACGAGGCTTAATAAAGTAGATGTTCCGCAAGA 240			
Db 181	AAAAGTTTCCCTTGGGATCGTTCGACGAGGCTTAATAAAGTAGATGTTCCGCAAGA 240			
QY 241	GGAGAGTAGAAGACCTCTAAACCCACCAATAGTCGGGTATGTAAGGATAGAAACGCTCTATA 300			

```

Db 241 GGAGGTAGAAACCTGAAACCCGACAAATAGCCGGTATGGAAGGATAGAAACGCTCTAA 300
Qy 301 TTCCACCATTCGACAGCGTTGGTCACGACCTCACTAAAGCCTAAACCGGATAGAG 360
Db 301 TTCCACCATTCGACAGCGTTGGTCACGACCTCACTAAAGCCTAAACCGGATAGAG 360
Qy 361 ACAGAGATGGAAGCCTAATAGAGAAACCTCGAGAGATAGAACCGTCATACCTAAAGTGAC 420
Db 361 ACAGAGATGGAAGCCTAATAGAGAAACCTCGAGAGATAGAACCGTCATACCTAAAGTGAC 420
Qy 421 AGATCACAGTGTCTAGCTAACGCCGACAAATAGCCCCAGCAGAGAAAGGTGAAAAACCA 480
Db 421 AGATCACAGTGTCTAGCTAACGCCGACAAATAGCCCCAGCAGAGAAAGGTGAAAAACCA 480
Qy 481 AATAGGACTCATCTAGCTAGCAGAGAGAGTACGGATATAGAAATCGTGAACGATCAAC 540
Db 481 AATAGGACTCATCTAGCTAGCAGAGAGAGTACGGATATAGAAATCGTGAACGATCAAC 540
Qy 541 TAATAGTAAGCAGTATTCCTGTTATATCGAGCATAGTTTTTATTTGACTTTTTAAACCCAG 600
Db 541 TAATAGTAAGCAGTATTCCTGTTATATCGAGCATAGTTTTTATTTGACTTTTTAAACCCAG 600
Qy 601 GGAACCCCTAACTGGATGGGTAGTTCTAGGATTTTTTTATT 642
Db 601 GGAACCCCTAACTGGATGGGTAGTTCTAGGATTTTTTTATT 642

RESULT 2
US-10-068-080-12
; Sequence 12, Application US/10068080
; Patent No. US20020115591A1
; GENERAL INFORMATION:
; APPLICANT: Frits, Christian
; APPLICANT: Youngman, Philip
; TITLE OF INVENTION: USE OF YNES, ESSENTIAL BACTERIAL GENES AND POLYPEPTIDES
; FILE REFERENCE: 07334/090001
; CURRENT APPLICATION NUMBER: US/10/068,080
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/070,116
; PRIOR FILING DATE: 1997-12-31
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 12
; LENGTH: 582
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-10-068-080-12

Query Match 15.6%; Score 100; DB 12; Length 582;
Best Local Similarity 52.9%; Pred. No. 3e-20;
Matches 239; Conservative 0; Mismatches 210; Indels 3; Gaps 1;

Qy 1 TACTAATGTTATCAAAATAATTAGGATCGGATAGACGACCCCAAGCTAAGGTAGACACGAG 60
Db 1 TACAATTAACGAAATAACTAATAAACCAGGATGAACATATCCGTCGTAAGGTAGACCCGAAT 60
Qy 61 ACCTAACTGTCATAGAAAGTTAGTATAGTCGCTCGTACCAAGACCATTTGTGACCT 120
Db 61 TAACACCCGTTTCGAACGGTTTCCTTAATATAGCCCTCGTCCGCTCGCGTGAATCCG 120
Qy 121 TGCCTGGTTGTGGAAGGGTAAATCCATCTCTTCGACATACCGCTTGGAAACACTAAGT 180
Db 121 CGATGGTTAGGTAGGATGAACCCACATTTTCGACCAAGCCAGCAGTATCGGCCTCTA 180
Qy 181 AAAAGTTTCTCGGATCGTTGCGAGAGGCTAATAAAGATGATGTTCCGCCAAGA 240
Db 181 TAAACTTTCTCGTACCGCTTGACGTAACGGAAAGAGTACGTACA---ACTATAAGTG 237
Qy 241 GGAGGTAGAAACCTGAAACCCGACAAATAGCCGGTATGGAAGGATAGAAACGCTCTAA 300
Db 238 GGCGAAGAACCTCTCAGAAACCGCAAAATCCCGGTGCACAAAGGGTAGAAGCGGTTTAA 297

```

```

Qy 301 TTTCACCATTCGACAGCGTTGGTCACGACCTCACTAAAGCCTAAACCGGATAGAG 360
Db 298 TTTCACCATTCGCGCACCGCTGTAGTCTCCGCAAAACGATAAAATCGTGGGACAAT 357
Qy 361 ACAGAGATGGAAGCCTAATAGAGAAACCTCGAGAGATAGAACCGTCATACCTAAAGTGAC 420
Db 358 AAATAGTGTCTACCAACGCCATAGAAAGTAAAGAAATATGAACCTGATTTAAACAAGAGAG 417
Qy 421 AGATCACAGTGTCTAGCTAACGCCGACCAATA 452
Db 418 AGTAGTACATTTGTCCCTAGATATGACAAATA 449

RESULT 3
US-09-790-988-1/c
; Sequence 1, Application US/09790988
; Patent No. US20020127687A1
; GENERAL INFORMATION:
; APPLICANT: SHIGENOBU, SHUJI
; APPLICANT: WATANABE, HIDEMI
; APPLICANT: HATTORI, MASAHIRO
; APPLICANT: SAKAKI, YOSHIYUKI
; TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
; FILE REFERENCE: 081356/0159
; CURRENT APPLICATION NUMBER: US/09/790,988
; CURRENT FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: JP2000-107160
; PRIOR FILING DATE: 2000-04-07
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 640681
; TYPE: DNA
; ORGANISM: Buchnera sp.
US-09-790-988-1

Query Match 5.5%; Score 35.6; DB 10; Length 640681;
Best Local Similarity 54.6%; Pred. No. 8.8;
Matches 71; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

Qy 471 TGAACCAACCAATAGGACTCATTTGATCTAGAGAGAGAGTAGCCGATATAGANTCGTGA 530
Db 291473 TGAACATATATATACGACCCCAATTCATTTGTAAGAAATTCATAAATGTA 291414
Qy 531 ACATCAACTATATAGTAAAGCAGTATTCCTGTTATATCGAGCATAGTTTATTTTACT 590
Db 291413 AAAATTAATTAATGATAATTCATTTATGAATAATACATAGATAATTTACTTTT 291354
Qy 591 TTTAAACCCAG 600
Db 291353 TTAATAATTAG 291344

RESULT 4
US-09-835-232-6/c
; Sequence 6, Application US/09835232
; Patent No. US20020098489A1
; GENERAL INFORMATION:
; APPLICANT: Leder, Philip
; APPLICANT: Leader, Benjamin
; TITLE OF INVENTION: FORMIN-2 NUCLEIC ACIDS AND POLYPEPTIDES
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 00383/052002
; CURRENT APPLICATION NUMBER: US/09/835,232
; CURRENT FILING DATE: 2001-04-12
; PRIOR APPLICATION NUMBER: US 60/196,811
; PRIOR FILING DATE: 2000-04-13
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 180216
; TYPE: DNA
; ORGANISM: Homo sapiens

```

```
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(180216)
; OTHER INFORMATION: n = A,T,C or G
US-09-835-232-6

Query Match          5.3%; Score 34.2; DB 10; Length 180216;
Best Local Similarity 56.8%; Pred. No. 13;
Matches 63; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 520 TAGAATCGTGAACGATCAAACTAATAGTAGACGAGTATTCCTGTTATATCGAGCATAGTTT 579
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 88856 TAGAATGGCAGCTGACTCAACCCATTAGTACTATATAGCCATATGGGTACAGTTATTATTA 88797
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 580 TTATTGTTGACTTTAAACACGAGGAAACCCCTAACTTGGATTGGGTAGTTCTTA 630
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 88796 TTATCATCATTTAAAAATGAGGTAGGCCCTCACGTCAGTGTCTTTGATCTA 88746
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 5
US-09-938-842A-2917/c
; Sequence 2917, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIPT1300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 2917
; LENGTH: 1058
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-2917

Query Match          5.2%; Score 33.6; DB 9; Length 1058;
Best Local Similarity 55.0%; Pred. No. 2;
Matches 66; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

QY 464 AGAAGGTGAAAAACCAATAGGACTCATTTGATCTAGAGAGAGAGTACCGATAATAGA 523
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 150 ACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 91
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 524 ATCGTGAACGATCAAACTAATAGTAGACGAGTATTCCTGTTATATCGAGCATAGTTTAT 583
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 90 AGCCTGAATTTAAAAATATTCATCATTTATTTATTCATTTCGACCAACCTTTATTATAT 31
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 6
US-09-864-761-33009
; Sequence 33009, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeonica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
```

```
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 33009
; LENGTH: 236
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC008166.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.3
; OTHER INFORMATION: NT HIT: ALI63246.2, EVALUE 6.00e-41
; OTHER INFORMATION: EST_HUMAN HIT: AI436093.1, EVALUE 4.00e-43
; OTHER INFORMATION: SWISSPROT HIT: P43679, EVALUE 3.90e-01
US-09-864-761-33009

Query Match          5.2%; Score 33.4; DB 10; Length 236;
Best Local Similarity 51.7%; Pred. No. 1.2;
Matches 76; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 430 TGTCTGCTAGTAAAGCCGACAAATAGCCCAAGACGAGAGAGGTGAAAACCAAAATAGGAC 489
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 58 TGTCAATCTAGCGTGGGACCAACAGTCACTCACTCACTCACTCACTCACTCACTCACTCACT 117
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 490 TCATTGATCTAGAGAGAGAGTACGATATAGTAATCGTGAACGATCAAACTAATAGTAA 549
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 118 TAATTGGATCCAGCAGCAGGAGCAAGTGGTGCCACTGAACCATCAAAATCAAGGTGA 177
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 550 GCAATATTCTCTTATATCGAGCATAG 576
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 178 GCATAGCTACTGTGAATAAACAGCAGAG 204
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 7
US-09-864-761-16513
; Sequence 16513, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
```



```
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6498
; LENGTH: 1155
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1155)
US-09-815-242-6498

Query Match          5.1%; Score 33; DB 10; Length 1155;
Best Local Similarity 49.7%; Pred. No. 3.2;
Matches 84; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 14 AAATAATTAGATCGGATAGACAGCAACCAAGCTAAGGTAGACACCACTAACTGTTTC 73
   || || || || || || || || || || || || || || || || || || || ||
DB 779 AATGGAATTTTGTATGAACAAGAACATGCAGCCTATATTGAAAAAATTATTCAACGTTTCC 838
   || || || || || || || || || || || || || || || || || || || ||
QY 74 ATAGAAAGTTTAGTTAGATCGCTCGTACCAAGACCACTTGACCTTGCTGGTTGTGGA 133
   || || || || || || || || || || || || || || || || || || || ||
DB 839 AAATAAATATATTTCAGATCTATTACACGTGTAGCAGCAGCAACCAATCAGAAAAATTAG 898
   || || || || || || || || || || || || || || || || || || || ||
QY 134 AGCGTAAATCCATTCTTTTCGACCATACCTGTTGGAACACTAACTGAA 182
   || || || || || || || || || || || || || || || || || || || ||
DB 899 GTGCGCAAGACGGTTTATTTCGACCAATCCGTGAATTACAGGAACGCAA 947
   || || || || || || || || || || || || || || || || || || || ||

RESULT 10
US-09-452-599-2
; Sequence 2, Application US/09452599
; Patent No. US2002055101A1
; GENERAL INFORMATION:
; APPLICANT: Bergeron, Michel G.
; APPLICANT: Ouellette, Marc
; APPLICANT: Roy, Paul H.
; TITLE OF INVENTION: Specific and Universal Probes and Amplification Primers
; TITLE OF INVENTION: to Rapidly Detect and Identify Common Bacterial
; TITLE OF INVENTION: Pathogens and Antibiotic Resistance Genes from Clinical
; TITLE OF INVENTION: Specimens for Routine Diagnosis in Micro
; FILE REFERENCE: 12287.31
; CURRENT APPLICATION NUMBER: US/09/452,599
; CURRENT FILING DATE: 1999-12-01
; PRIOR APPLICATION NUMBER: 08/526,840
; PRIOR FILING DATE: 1995-09-11
; PRIOR APPLICATION NUMBER: 08/304,732
; PRIOR FILING DATE: 1994-09-12
; NUMBER OF SEQ ID NOS: 177
; SOFTWARE: PatentIn Ver. 2.1
```

```
; SEQ ID NO 2
; LENGTH: 2275
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
US-09-452-599-2

Query Match          5.1%; Score 33; DB 10; Length 2275;
Best Local Similarity 49.7%; Pred. No. 4.3;
Matches 84; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 14 AAATAATTAGATCGGATAGACAGCAACCAAGCTAAGGTAGACACCACTAACTGTTTC 73
   || || || || || || || || || || || || || || || || || || || ||
DB 1612 AATGGAATTTTGTATGAACAAGAACATGCAGCCTATATTGAAAAAATTATTCAACGTTTCC 1671
   || || || || || || || || || || || || || || || || || || || ||
QY 74 ATAGAAAGTTTAGTTAGATCGCTCGTACCAAGACCACTTGACCTTGCTGGTTGTGGA 133
   || || || || || || || || || || || || || || || || || || || ||
DB 1672 AAATAAATATATTTCAGATCTATTACACGTGTAGCAGCAGCAACCAATCAGAAAAATTAG 1731
   || || || || || || || || || || || || || || || || || || || ||
QY 134 AGCGTAAATCCATTCTTTTCGACCATACCTGTTGGAACACTAACTGAA 182
   || || || || || || || || || || || || || || || || || || || ||
DB 1732 GTGCGCAAGACGGTTTATTTCGACCAATCCGTGAATTACAGGAACGCAA 1780
   || || || || || || || || || || || || || || || || || || || ||

RESULT 11
US-09-785-474-1/c
; Sequence 1, Application US/09785474
; Patent No. US20010012626A1
; GENERAL INFORMATION:
; APPLICANT: TANZI, RUDOLPH
; APPLICANT: WASCO, WILMA
; TITLE OF INVENTION: Genetic Alterations Related To Familial
; TITLE OF INVENTION: Alzheimer's Disease
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA: US/09/785,474
; APPLICATION NUMBER: 08/706,344
; FILING DATE: 30-AUG-1996
; APPLICATION NUMBER: 60/003,054
; FILING DATE: 31-AUG-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: KIM, JUDITH U.
; REGISTRATION NUMBER: 40,679
; REFERENCE/DOCKET NUMBER: 0609.4180002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2765 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 249..1649
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-785-474-1
```

Query Match 5.1%; Score 33; DB 10; Length 2765;
Best Local Similarity 55.6%; Pred. No. 4.7;
Matches 85; Conservative 0; Mismatches 65; Indels 3; Gaps 1;

QY 111 ATTGTGACCTTGGTGGTGAAGCGGTAAATCCATTCTTTGACCATACCGTTGGAA 170
|| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 1756 ATCTTGACTTTGTTAGATGGACACAGAAATCACCCTTTGCTCCCGAGATTGGTT 1697

QY 171 ACACCTAACTGAAAGTTCCTTGGGATCGTTGCGAGCA--GCCTAATAAAGTAGA 227
|| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 1696 ATCTTGACTTTGTTAGATGGACACAGAAATCACCCTTTGCTCCCGAGATTGGTT 1697

QY 171 ACACCTAACTGAAAGTTCCTTGGGATCGTTGCGAGCA--GCCTAATAAAGTAGA 227
|| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 1696 ATCTTGACTTTGTTAGATGGACACAGAAATCACCCTTTGCTCCCGAGATTGGTT 1697

QY 228 TGTTCGCCAAGAGGAGTAGAAACCTGAAAA 260
|| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 1636 TGAATGCTAATTGGTCCATATAAAGGCTGTACA 1604

RESULT 12

US-09-785-474-3/c
; Sequence 3, Application US/09785474
; Patent No. US20010012626A1
; GENERAL INFORMATION:
; APPLICANT: TANZI, RUDOLPH
; WASCO, WILMA
; TITLE OF INVENTION: Genetic Alterations Related To Familial
; Alzheimer's Disease
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/785,474
; FILING DATE: 20-Feb-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/706,344
; FILING DATE: 30-AUG-1996
; APPLICATION NUMBER: 60/003,054
; FILING DATE: 31-AUG-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: KIM, JUDITH U.
; REGISTRATION NUMBER: 40,679
; REFERENCE/DOCKET NUMBER: 0609.4180002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2765 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 249..1649
; SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-785-474-3

Query Match 5.1%; Score 33; DB 10; Length 2765;
Best Local Similarity 55.6%; Pred. No. 4.7;
Matches 85; Conservative 0; Mismatches 65; Indels 3; Gaps 1;

QY 111 ATTGTGACCTTGGTGGTGAAGCGGTAAATCCATTCTTTGACCATACCGTTGGAA 170
|| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 1756 ATCTTGACTTTGTTAGATGGACACAGAAATCACCCTTTGCTCCCGAGATTGGTT 1697

QY 171 ACACCTAACTGAAAGTTCCTTGGGATCGTTGCGAGCA--GCCTAATAAAGTAGA 227
|| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 1696 ATCTTGACTTTGTTAGATGGACACAGAAATCACCCTTTGCTCCCGAGATTGGTT 1697

QY 228 TGTTCGCCAAGAGGAGTAGAAACCTGAAAA 260
|| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 1636 TGAATGCTAATTGGTCCATATAAAGGCTGTACA 1604

RESULT 13

US-09-785-474-27/c
; Sequence 27, Application US/09785474
; Patent No. US20010012626A1
; GENERAL INFORMATION:
; APPLICANT: TANZI, RUDOLPH
; WASCO, WILMA
; TITLE OF INVENTION: Genetic Alterations Related To Familial
; Alzheimer's Disease
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/785,474
; FILING DATE: 20-Feb-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/706,344
; FILING DATE: 30-AUG-1996
; APPLICATION NUMBER: 60/003,054
; FILING DATE: 31-AUG-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: KIM, JUDITH U.
; REGISTRATION NUMBER: 40,679
; REFERENCE/DOCKET NUMBER: 0609.4180002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2765 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 249..1649
; SEQUENCE DESCRIPTION: SEQ ID NO: 27:
US-09-785-474-27

Query Match 5.1%; Score 33; DB 10; Length 2765;
Best Local Similarity 55.6%; Pred. No. 4.7;
Matches 85; Conservative 0; Mismatches 65; Indels 3; Gaps 1;

QY 111 ATTGTGACCTTGGTGGTGAAGCGGTAAATCCATTCTTTGACCATACCGTTGGAA 170
|| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 1756 ATCTTGACTTTGTTAGATGGACACAGAAATCACCCTTTGCTCCCGAGATTGGTT 1697

QY 171 ACACCTAACTGAAAGTTCCTTGGGATCGTTGCGAGCA--GCCTAATAAAGTAGA 227

: 710.61 secs

Search completed: December 26, 2002, 00:46:05
Job time : 710.61 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 24, 2002, 22:13:55 ; Search time 10196.6 seconds
(without alignments)
1661.118 Million cell updates/sec

Title: US-10-068-080-12

Perfect score: 582

Sequence: 1 tacaattacgaataacta.....gatttcatttaccataatt 582

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

1: gb.ba:*
2: gb.htg:*
3: gb.in:*
4: gb.om:*
5: gb.ov:*
6: gb.pat:*
7: gb.ph:*
8: gb.pl:*
9: gb.pr:*
10: gb.ro:*
11: gb.sts:*
12: gb.sy:*
13: gb.un:*
14: gb.vi:*
15: em.ba:*
16: em.fun:*
17: em.hum:*
18: em.in:*
19: em.mu:*
20: em.om:*
21: em.or:*
22: em.ov:*
23: em.pat:*
24: em.ph:*
25: em.pl:*
26: em.ro:*
27: em.sts:*
28: em.un:*
29: em.vi:*
30: em.htg_hum:*
31: em.htg_inv:*
32: em.htg_other:*
33: em.htg_mus:*
34: em.htg_pln:*
35: em.htg_rod:*
36: em.htg_mam:*
37: em.htg_vrt:*
38: em_sy:*
39: em.htgo_hum:*
40: em.htgo_mus:*
41: em.htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	49	8.4	4574	9	HSM803437	AL832130 Homo sapi
2	48	8.2	7218	6	I66494	I66494 Sequence 14
3	47.6	8.2	152103	2	AC012130	AC012130 Homo sapi
4	47.6	8.2	186609	9	AC083872	AC083872 Homo sapi
5	46.6	8.0	12025	6	AX346200	AX346200 Sequence
6	46	7.9	24175	2	AC115594	AC115594 Dictyoste
7	45.4	7.8	183345	2	AC120550	AC120550 Mus muscu
8	45.4	7.8	203422	2	AC107857	AC107857 Mus muscu
9	45.2	7.8	17480	3	AC114258	AC114258 Dictyoste
10	44.8	7.7	105470	2	AC116306	AC116306 Dictyoste
11	44.4	7.6	6641	6	AX281294	AX281294 Sequence
12	44.4	7.6	6641	6	AX345217	AX345217 Sequence
13	44.4	7.6	120571	9	CNS01073	AL132633 Human chr
14	44.4	7.6	172311	2	AC021729	AC021729 Homo sapi
15	44.4	7.6	218956	9	CNS00008	AL049831 Human chr
16	44.2	7.6	124820	2	AC117073	AC117073 Dictyoste
17	44.2	7.6	168543	2	AL845301	AL845301 Danio rer
18	44	7.6	58894	9	AC093277	AC093277 Homo sapi
19	44	7.6	170392	9	AC060765	AC060765 Homo sapi
20	44	7.6	180697	9	AC114977	AC114977 Homo sapi
21	43.8	7.5	152409	2	PFMAL1P1	AL031744 Plasmodiu
22	43.6	7.5	100920	9	AC020921	AC020921 Homo sapi
23	43.6	7.5	167872	2	AC007554	AC007554 Homo sapi
24	43.6	7.5	251664	2	AC008908	AC008908 Homo sapi
25	43.4	7.5	86827	3	PFMAL3P5	AL034556 Plasmodiu
26	43	7.4	158376	9	AC099059	AC099059 Homo sapi
27	43	7.4	194281	2	AC078812	AC078812 Homo sapi
28	42.8	7.4	69048	2	AC011826	AC011826 Homo sapi
29	42.6	7.3	131201	2	AC130820	AC130820 Mus muscu
30	42.6	7.3	177604	2	AC013342	AC013342 Homo sapi
31	42.6	7.3	183338	2	AC083775	AC083775 Homo sapi
32	42.6	7.3	188014	9	AC093744	AC093744 Homo sapi
33	42.2	7.3	1852	9	BC028121	BC028121 Homo sapi
34	42.2	7.3	57835	2	AC110290	AC110290 Homo sapi
35	42	7.2	77835	2	PFMAL13P2_3	Continuation (4 of
36	42	7.2	168469	9	AC090592	AC090592 Homo sapi
37	42	7.2	168482	2	AC025235	AC025235 Homo sapi
38	42	7.2	177628	2	AL772143	AL772143 Danio rer
39	41.3	7.2	2863	3	EHAVAMER	X75436 E.histolyti
40	41.3	7.2	2913	3	ENHRRSA	M95498 Entamoeba h
41	41.3	7.2	3188	3	ENHRRSA	L36807 Entamoeba h
42	41.3	7.2	30726	2	AC117269	AC117269 Dictyoste
43	41.3	7.2	65691	3	PFMAL3P1	Z97348 Plasmodium
44	41.3	7.2	98734	2	PFMAL1P2	AL031745 Plasmodiu
45	41.3	7.2	106216	9	AC061997	AC061997 Homo sapi

ALIGNMENTS

RESULT 1
HSM803437
LOCUS Homo sapiens mRNA; cDNA DKF2p666H1210 (from clone DKF2p666H1210).
DEFINITION 4574 bp mRNA linear PRI 12-JUL-2002
ACCESSION AL832130
VERSION AL832130.1 GI:21732673
KEYWORDS human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 4574)
AUTHORS Wambutt,R., Heubner,D., Mewes,H.W., Weil,B. and Wiemann,S.
TITLE Direct Submission
JOURNAL Submitted (09-JUL-2002) 1, D-85764 Neuherberg, GERMANY

COMMENT
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by AGWA (Berlin/Germany) within the cDNA sequencing consortium of the German Genome Project.
 This clone (DKFZp686h1210) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further information about the clone and the sequencing project is available at http://mips.gsf.de/proj/cDNA/.

FEATURES
 Location/Qualifiers

source
 1. .4574
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="DKFZp686h1210"
 /tissue_type="cDNA-collection"
 /clone_lib="686 (synonym: hlcc3). Vector pSport1_Sfi; host DH10B; sites SfiIA + SfiIB"
 /dev_stage="adult"
 /dev_stage="adult"
 polyA_signal
 polyA_site
 4394. .4399
 BASE COUNT 1349 a 1085 c 1074 g 1066 t
 ORIGIN

Query Match 8.4%; Score 49; DB 9; Length 4574;
 Best Local Similarity 47.5%; Pred. No. 0.038;
 Matches 145; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

Qy 213 AAAGAGTACGTACACTATAAGTGGCGGAAGACGCTCTCAGAAACGCCAAATCCGGT 272
 Db 4270 AAAAAAAAAAAAAAAAAAATCCCTGAATGATTAGACATCAGCGCTAAAAAACTAC 4329
 Qy 273 GCACAAAGGCGTGAAGCGGTTTAAATTTCCGCCATTTCCGCCCGCTGTAGTCTCCGCA 332
 Db 4330 ATTATAGCTAGAGTTGTTATATATGCAATATTTCTGCTCTTCTTTGTTCTGTTA 4389
 Qy 333 AAACGATAAATCGTGGGACATAAATAGTCTACCAACGCCATGAAGTAGAGAAAAA 392
 Db 4390 AAACATAAATGCAATGTTGATATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4449
 Qy 393 TATGAAGTATTTAAACAAGAGAGTACGTACATTTGCTCCCTAGATGACAATATAT 452
 Db 4450 AA 4509
 Qy 453 ATCAAGAACAGCTACTATGCTATAATAACTAACAGCATGGGAGCGTGTATAAACA 512
 Db 4510 AA 4569
 Qy 513 CTATA 517
 Db 4570 AAAAA 4574

RESULT 2
 I66494/c 166494
 LOCUS
 DEFINITION Sequence 14 from patent US 5670367.
 ACCESSION 166494
 VERSION 166494.1 GI:2724471
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 7218)
 AUTHORS Dörner, F., Scheiflinger, F. and Falkner, F. Gunter.
 TITLE Recombinant fowlpox virus
 JOURNAL Patent: US 5670367-A 14 23-SEP-1997;
 FEATURES Location/Qualifiers
 source
 1. .7218
 /organism="unknown"
 BASE COUNT 1944 a 1491 c 1486 g 1929 t 368 others
 ORIGIN

Query Match 8.2%; Score 48; DB 6; Length 7218;

Best Local Similarity 2.7%; Pred. No. 0.069;
 Matches 9; Conservative 194; Mismatches 129; Indels 0; Gaps 0;

Qy 203 GACGTAAACGAAAGAGTACGTACAACTATAAGTGGCGGCAAGAACGCTCTCAGAAACGCC 262
 Db 1382 RRR 1323
 Qy 263 AAAATCCGTCGACAAAGGTTAGAACGGTAAATTTCCGCCATTTCCGCCACCGCTGTA 322
 Db 1322 RRR 1263
 Qy 323 GTCCTCCGCAAAACGATAAAATCGTGGGACAAATAAATAGTCTACCAACGCCATAAGA 382
 Db 1262 RRR 1203
 Qy 383 ACTAGAAAATATGAATGATTAAACAAGAGAGAGTAGTACATATGTCCTACATAT 442
 Db 1202 RRR 1143
 Qy 443 GACAATATATCAAAAGAACAGGTACTATGCAATAAATAACTAACAGCAATGGGACGAGT 502
 Db 1142 RRR 1083
 Qy 503 GATAAAACACTATATGCTGTGGTGGCTGTTG 534
 Db 1082 RRR 1051

RESULT 3
 AC012130 152103 bp DNA linear HTG 30-MAR-2000
 LOCUS Homo sapiens clone RP11-4G15, WORKING DRAFT SEQUENCE, 8 unordered
 DEFINITION pieces
 AC012130
 AC012130.2 GI:7341840
 VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
 KEYWORDS Homo sapiens.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
 Baldwin,J., Barna,N., Becker,K., Boguslavsky,L., Bouckhalter,B.,
 Brown,A., Castle,A., Collins,S., Collins,S., Collins,A.,
 Cooke,P., DeArrellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
 Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
 Galagan,J., Gardyna,S., Grant,G., Hags,B., Heford,A., Horton,L.,
 Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
 Lehoczy,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
 McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Melnick,J.,
 Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
 Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
 Stange-Thomann,N., Stojanovic,N., Subramanian,A., Takamas,J.,
 Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
 Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (20-OCT-1999) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Mar 30, 2000 this sequence version replaced gi:6087954.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 ----- Project name: L2823

Qy	441	ATGACAATATATATCAAGAAACAGGTA	CTATGCA	TAAATAACAC	CAATGGGACA	500
Db	54204	TTGCCTTTAGATACAAAGTCAGGGA	TTTCTG	TTCAA	TATTTACTG	TACATATACA 54263
Qy	501	GTGATAAAAACACTATATGCTGTGGC	TGCGTCTG	CTTAATTTG	CGTTAATAGT	TATTTTGTCT 560
Db	54264	GTCTTCGACACAAACAGCTTTGAGA	AGGCTTGG	TATTTCC	ATTATAA	TAAATTTGAGA 54323
Qy	561	TGGATTTTCATTTTACCAA	578			
Db	54324	CGTAGGGCATT	TTTTCAAA	54341		

RESULT 4	AC083872	186609 bp	DNA	linear	PRI 31-JAN-2
LOCUS	AC083872				
DEFINITION	Homo sapiens chromosome 7 clone RP11-639H21, complete sequence.				
ACCESSION	AC083872				
VERSION	AC083872.2	GI:12621403			
KEYWORDS	ETG.				
SOURCE	Homo sapiens.				

ORGANISM	Homo sapiens.
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS	1 (bases 1 to 186609) Kaul, R.K., Zhou, Y., James, R.A., Raymond, C., Haugen, E.D. and Olson, M.V.
TITLE	Large-scale Mapping and Sequencing of Human Chromosome 7
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 186609)
AUTHORS	Kaul, R., Olson, M.V., Raymond, C., Clendenning, J. and Haugen, E.
TITLE	Direct Submission
JOURNAL	Submitted (04-OCT-2000) Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA
REFERENCE	3 (bases 1 to 186609)

Location/Qualifiers
1. .152103

```

TITLE
JOURNAL
-----
Submitted (31-JAN-2001) Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA
On Jan 31, 2001 this sequence version replaced gi:10567937.
-----
COMMENT
-----
Center: University of Washington Genome Center
Center Code: UWGC
Web site: http://www.genome.washington.edu
Contact: uwgchits@u.washington.edu
-----
Project Information
-----
Center project name: HsaChr7
Center clone name: Rpl1-639H21 (djs716)
-----
Summary Statistics
-----
Sequencing vector: plasmid; X52328; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 186209 bases at least Q40
Consensus quality: 186563 bases at least Q30
Consensus quality: 186599 bases at least Q20
Insert size: 299800; 15.1% error; agarose-ftp
Insert size: 186609; sum-of-contigs
Quality coverage: 10.09x in Q20 bases; agarose-ftp
Quality coverage: 11.35x in Q20 bases; sum-of-contigs

```

Overlapping Sequences:
 5': RP11-344L16 (WGC:djs322) AC018646, 13611-bp clone over
 3': RP11-343M10 (WGC:djs732)

Sequence Quality Assessment:
This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.

All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp. Base-by-base quality values are not generally visible from the

Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part

of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted:
all regions were either double-stranded or sequenced with an
alternate chemistry or covered by high quality data (i.e., Phred
quality >= 30); an attempt was made to resolve all sequencing
problems, such as compressions and repeats; all regions were
covered by at least one plasmid subclone or more than one M13
subclone; and the assembly was confirmed by restriction digest.

Sequence Validation:

This sequence has been validated by Multiple Complete Digest
fingerprinting. Comparison of the experimentally derived digest
fragments with sequence-predicted fragments is given below.
The electronically-digested sequence consists of both insert and
vector, in order to accurately represent the entire circular BAC.
Small fragments below a variable cutoff (approximately 400-800 bp)
are not resolved in the fingerprint and hence do not appear
in the table. There are no significant remaining discrepancies
between the experimental and predicted values. Uniquely ordered
fragments are separated by dashed lines.

EcRI BglII HindIII

SeqDerMap	FngrPrint	SeqDerMap	FngrPrint	SeqDerMap	FngrPrint
8696	8908	9428	9477	1091	1108
6	<800	2067	2034	6382	6457
1282	1295	5146	5237	512	<800
5875	5795	214	<800	449	<800
864	875	6664	6551	2332	2341
3210	3260	2112	2139	3406	3432
547	<800	558	<800	4842	4741
2719	2686	2175	2139	341	<800
12	<800	2550	2568	6269	6086
3603	3651	2441	2485	6163	6086
4317	4277	8416	8390	2196	2194
579	<800	4875	4728	171	<800
4039	4057	1489	1456	939	957
8217	8202	15165	15187	679	<800
1539	1508	3130	3275	4331	4364
371	<800	1521	1456	618	<800
16481	16182	6483	6345	5175	5188
695	<800	1151	1140	1062	1108
1217	1182	609	<800	1506	1461
729	721	4775	4728	4663	4741
7979	8202	3602	3532	2141	2194
1517	1508	282	<800	275	<800
304	<800	1257	1264	9412	9422

5107	6102	10	<800	1133	1108
11972	12030	2735	2848	1290	1261
893	875	3563	3532	5872	6086
828	875	11001	11223	1949	1916
1149	1182	4671	4728	1430	1461
5513	5513	1494	1456	2544	2595
8138	8202	5095	5079	2555	2595
131	<800	3991	3944	6162	6086
1346	1295	1450	1456	2058	2194
3683	3651	2231	2267	6346	6457
285	<800	1043	1031	41	<800
480	<800	11317	11223	3020	2972
10788	10771	4313	4310	56	<800
12229	12030	1819	1789	6425	6457
10595	10339	2613	2678	474	<800
1804	1800	5863	5777	4380	4364
19511	19785	1410	1456	2187	2194
6130	6102	4258	4111	3234	3266
5058	5030	235	<800	1460	1461
262	<800	1484	1456	5208	5188
4871	4887	209	<800	3224	3266
5728	5795	539	<800	161	<800
3009	3046	1298	1264	2313	2341
2982	3061	2904	2972		
6198	6120	3309	3266		
871	862	601	<800		
2286	2387	124	<800		
457	<800	14221	14186		
3725	3698	6344	6086		
432	<800	2218	2194		
2926	2968	4420	4364		
1590	1662	3743	3764		
70	<800	4425	4741		
333	<800	724	<800		
274	<800	247	<800		

Query Match 8.2%; Score 47.6; DB 9; Length 186609;
Best Local Similarity 52.5%; Pred. No. 0.088;
Matches 104; Conservative 0; Mismatches 94; Indels 0; Gaps 0;


```

/protein_id="AAL92303.1"
/db_xref="GI:19570003"
/translation="MAHLLQKDRFFLYQEPYIAGRCRGAAGVETTRLDIGTARN5
YDPGFERDDKSLTNTGDEDEDEKNGGGGNGRFRDEFEGSSDUDKLFKNGK
DYGDDKDAIDWIDSKMDSRRKKREANEQKIEKQRESRPKIQQHLDKQDLAT
VSDQWASLPDAGDLRRNRQKHEIYTPVPSLVERAKSENETYSLQTNNAATAPG
IGLSGDSNGSTGTDLTQVGSARKTVLDLKHQVSDNYSKTCVDPKGYLTDLRSKR
IASDAEVDGDIKKARLLFKKATTSNPKHSPSIAAAKLEVLAVGKIVDARKIIAOCREK
PTSEVEANENLQDPAKIVLAQAVRVPISVKIWLVAANLEKOLKMKKRYLRAL
EPIPTSVKLKEAVELEEDPARILGLRAVECPVDNIDMLALANLEYTEKAREVLNK
ARALPSSSEIIVISAQLESQGNNDVNNKIIKKAISKCSGVMYNNRDKWIAEAEK
SEKNQYVTCQAIYYETIGMIEDDRKRIWIDAEELSRGSIKNNAYIAHILSIF
PTKSVMLKVAQLEKHGKGTLEQDQTLKATQCPQVENLWLMYAKEWIKSGVAVAKR
EIIAOFKPNPGENIWAAVKIESEMNELRAARNLLKRAIEAGTERIMWKSAALLER
ELGDDQKLESLIEQGLQKYPNSFKLWLMKAOLEEQATANHONNONNONNONNO
NQNNNLQDILSIEIIRQYTKNTVKCPNNGSINGIAOASPKFKNQNFNRALLESQ
AKLNQDIDIFLEVRFEDSLGNKQOATILALGIAOASPKFKQWELAEIEMPRHSQ
RNCVDALNRNDNPVFTQVSKIFWMDGKLDKAKOQVQRTVTFPFEGDGNAYYAF
ILKPEQNDKILKCVGEAPNLEHGIWKVSKQVDNSRLKTDQILKLVSNLSKSLQ
QOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQO
join(10072..10162,10293..10570,10693..10980)
/note="ORF_ID:dd_01564"
/codon_start=1
/product="pre-mRNA splicing factor prpl."
/protein_id="AAL92304.1"
/db_xref="GI:19570004"
/translation="MNLFSFLGDLHLGSLMLILLFKIKNDKSCAGVLSKSLQILTFIVF
TAYLDLFTNVSLIYTFMKITIIAVSYTLHLIARKYKFTYDKDHDHTFKIVYLIASC
AIIISLTIGTIGIYSFLEILTFISYILESTIAILPOLILQRTGEVEALTSNYIVL
LGGYRAFPLFNWYIRITFYNWSGKIEMLSLGLQTLIYADFFYYIYAKSRMYGKKLVLPQ
.
complement(11445..11663)
/note="ORF_ID:dd_01565"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAL92305.1"
/db_xref="GI:19570005"
/translation="MTPYLKIIKSSVTLISFFYFIANTIIRTIQNVPTSHKIILVSLY
YLVSFLTRIFRFGSLPKIISTYIYKGF"
complement(11978..12772)
/note="ORF_ID:dd_01566"
/codon_start=1
/product="3-hydroxyacyl-coA dehydrogenase type II"
/protein_id="AAL92306.1"
/db_xref="GI:19570006"
/translation="MKINGKTFVVTGGASGLGLETSHRLSLCANIFIMDMNEENGRK
VVELGDKTRFMFSDVITLSDSVKLSLEHCKRPFKEHGVINCAGVAARVIRKDGQ
VHPLDLFTVVMVNLIGTFNVIRLVADIHNONPSKDGDEEEKEKGVFIMTASVAA
FEGQVGOSAYSASKGVGMPLMAREFATLIRINTIAPGTFTPMVEMLPQAAIKS
INESIPPSRMGRKPEFAFLCOHLIENTYINGEIVIRLDGALRLSKL"
complement(join(13202..13450,13663..13703,13785..13891,
14179..14183))
/note="ORF_ID:dd_01567"
/codon_start=1
/product="40S RIBOSOMAL PROTEIN S23"
/protein_id="AAL92308.1"
/db_xref="GI:19570008"
/translation="MCKPNGLKTKARKLNHRRVORVNDKLYKKAHFGTALKANPFQGA
SHASGVPNFAIKRCVQVQLKNGKIKITAFVNDGCLNIQENDKVLVLAGIGRSGHSV
GDIPGVRFKVVKSSVSLIAIRGIKDPNV"
16633..20067
/note="ORF_ID:dd_01599"
/codon_start=1
/product="INTERSECTIN LONG FORM"
/protein_id="AAL92307.1"
/db_xref="GI:19570007"
/translation="MKIKFRVYLITDEQSEENKEVAIQQVNVNDTNSIDSSNRDSL
NIEEOQKSSSLSSNNNTTTTATPTTATTTTATTTTATTTTATTTTDESETNSTTP
TSATSPDSMITPTKSGSNKLSNEDKLLKKOKKEKLEKLEKLEKLEKLEKLEKLEK
KDEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEKLEK
KEKEDTEQDKDOSTPKVOLTREPSYIFSKVNSNLSIIERSTSPVQIDSSPSKS
FIESNNNNSSNNANNTTDRNISSEDANIIIONNTNNIENNTNNNNNSGR
TSTPTPINSEKIEKSISSSDLSNITTTTTTTTGGGGLNTSPSTENINVGGGTDSCGTTT

```

```

TTNNNNNNNNLVDKTELSSGSSVSSPNGTKREVVIKIVTRAFQVHONVMKEE
GKNKRNWVINEIINTEKDYIADLNIIIVNFILTPLESKILIPDKDIIIFSNIOLLN
VYKELSDLLQKAPDEKPNHTVGTCTIYFFKYLKIYSGCANOTISADHIAKCSKAP
SRQFLEEQKASTECQCNLESFLIKPQVRLQKYPULLRELIIKNSCEHPDLANLEKA
YTEIQEVLVSVNRESKRAEIKHOKMYIKHEKLAQPEKFFVFTTRYLIRAEATCELTEE
KOKISGRMHYILFNDIIMRWQDKKSIRLETFLI IASTNVDKGEENRSGSCFTATNPS
PGTTRGRFFLVADSVBOKMFTDLDELIPHREYSRNADKVLDDHIKRSATITNPS
SNGISSDPSKSLHRTQTAAGALQOQOQSKPVHAIPSKPLPPPPPKSKSATITPTTT
TNTVTVIDEPTLVDASTPLPSOTPIAKPRTSTIKKTEFTNSNOTSSPOQOQAT
TQOAPPIRQORPLSVITPTTTPSPQVITQONENNNTVTKTRPLPKPGSGFTAPA
PSTLNNNTNSNTTTTTTTTINQPAKPIAPPKSLNKLQAAPNSARSASTFVPS
THSSLVPSVSKNNTNNITKDLSPSSANNNNNNNTDFTTPTTGRPRVMPLPFR
GGSAVVANNNNNDNRVPPANTTGSAAHPHSQTSGLNVNKKPIIPPRK"
complement(join(20510..20790,20881..21217))
/note="ORF_ID:dd_01560"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAL92309.1"
/db_xref="GI:19570009"
/translation="MGSGESKFNQIAQOSRLDHADLTLLKQFNAISKNGEIKKDDL
KRIIEKYGSDSLTALYINLDFRDGDKTINTEFALAYGLVYKNTLDDVDTSFHCH
DLUNGDNISGELURAVVMANKMKMEYIQNDKQVPLDKLTLPIEVSKINQEAADDLFQ
LDVNDGVSQKQEFIQIOLASSSPKLLKLTSLYVRDEALDFGFTK"
complement(join(23066..23121,23286..24122))
/note="ORF_ID:dd_01561"
/pseudo
/codon_start=1
BASE COUNT 9560 a 3026 c 2700 g 8889 t
ORIGIN
Query Match 7.9% Score 46; DB 2; Length 24175;
Best Local Similarity 51.5%; Pred. No. 0.23;
Matches 106; Conservative 0; Mismatches 100; Indels 0; Gaps 0;
QY 377 ATAGAGAGTAGAAAATATGAACCTGATTTAAACAAGACAGAGAGTAGTACAAATGTGCCCT 436
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 173 ATATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 232
QY 437 AGATATGACATATATATCAAGAACACAGCTACTATGCATATAATACTACAGCAATGGG 496
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 233 ATTAATAAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAG 292
QY 497 ACAGGTGATAAAACACATATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 556
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 293 AATAATAATAAAACACTTAATAATTCCTCGCGGTTTGTGATTTTGTGATTTTGTGATTTT 352
QY 557 GTCTTGATTTTCATTTTACCAATATT 582
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 353 TTTTATTTTATTTTGTGTTTATTTT 378
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
RESULT 7
AC120550 183345 bp DNA linear HTG 09-MAY-2002
LOCUS Mus musculus clone RP23-14218, WORKING DRAFT SEQUENCE, 27 ordered
DEFINITION pieces.
ACCESSION AC120550
VERSION AC120550.2 GI:20514889
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 183345)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,A., Allen,N.,
TITLE Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
JOURNAL Bouknight,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
REFERENCE Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
AUTHORS Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
2 (bases 1 to 183345)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
Bouknight,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,

```

Faró, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Larocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G., MacLean, C., MacDonald, P., Major, J., Marquis, N., Meneus, L., McCarthy, M., McEwan, P., McKernan, K., Melrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (08-MAY-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 183345)

TITLE

JOURNAL

REFERENCE

AUTHORS

Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faró, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Larocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G., MacLean, C., MacDonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Melrim, J., Meneus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (09-MAY-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 9, 2002 this sequence version replaced gi:20503162.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

TITLE

JOURNAL

COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information -----
Center project name: L20738
Center clone name: 142.I.8
----- Summary Statistics -----

Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 173947 bases at least Q40
Consensus quality: 175704 bases at least Q30
Consensus quality: 176645 bases at least Q20
Insert size: 176000; agarose-fp
Insert size: 180745; sum-of-contigs
Quality coverage: 7.9 in Q20 bases; agarose-fp
Quality coverage: 7.7 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 27 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces

* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

1 183: contig of 183 bp in length
184 283: gap of 100 bp
284 982: contig of 699 bp in length
983 1082: gap of 100 bp
1083 1790: contig of 708 bp in length
1791 1890: gap of 100 bp
1891 2497: contig of 607 bp in length
2498 2597: gap of 100 bp
2598 3293: contig of 696 bp in length
3294 3393: gap of 100 bp
3394 4051: contig of 658 bp in length
4052 4151: gap of 100 bp
4152 4811: contig of 660 bp in length
4812 4911: gap of 100 bp
4912 34646: contig of 29735 bp in length
34647 34746: gap of 100 bp
34747 35900: contig of 1154 bp in length
35901 36000: gap of 100 bp
36001 37699: contig of 1699 bp in length
37000 37799: gap of 100 bp
37800 38952: contig of 1153 bp in length
38953 39052: gap of 100 bp
39053 41334: contig of 2282 bp in length
41335 41434: gap of 100 bp
41435 43398: contig of 1964 bp in length
43399 43498: gap of 100 bp
43499 45252: contig of 1754 bp in length
45253 45352: gap of 100 bp
45353 48902: contig of 3550 bp in length
48903 49002: gap of 100 bp
49003 52020: contig of 3018 bp in length
52021 52120: gap of 100 bp
52121 57368: contig of 5248 bp in length
57369 57468: gap of 100 bp
57469 65330: contig of 7862 bp in length
65331 65430: gap of 100 bp
65431 74541: contig of 9111 bp in length
74542 74641: gap of 100 bp
74642 85758: contig of 11117 bp in length
85759 85858: gap of 100 bp
85859 96006: contig of 10148 bp in length
96007 96106: gap of 100 bp
96107 108889: contig of 12783 bp in length
108890 108989: gap of 100 bp
108990 119894: contig of 10905 bp in length
119895 119994: gap of 100 bp
119995 134363: contig of 14369 bp in length
134364 134463: gap of 100 bp
134464 154899: contig of 20436 bp in length
154900 154999: gap of 100 bp
155000 183236: contig of 28237 bp in length
183237 183336: gap of 100 bp
183337 183345: contig of 9 bp in length.

FEATURES

source

1. .183345
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP23-142I8"
/clone_lib="RPCI-23 Female Mouse BAC"
1. .183
/note="assembly_fragment"
284. .982
/note="assembly_fragment"
1083. .1790
/note="assembly_fragment"
1891. .2497
/note="assembly_fragment"

misc_feature

misc_feature

misc_feature

misc_feature

```

misc_feature      2598..3293
                    /note="assembly_fragment"
misc_feature      3394..4051
                    /note="assembly_fragment"
misc_feature      4152..4811
                    /note="assembly_fragment"
misc_feature      4912..34646
                    /note="assembly_fragment"
misc_feature      34747..35900
                    /note="assembly_fragment"
misc_feature      36001..37699
                    /note="assembly_fragment"
misc_feature      37800..38952
                    /note="assembly_fragment"
misc_feature      39053..41334
                    /note="assembly_fragment"
misc_feature      41435..43398
                    /note="assembly_fragment"
misc_feature      43499..45252
                    /note="assembly_fragment"
misc_feature      45353..48902
                    /note="assembly_fragment"
misc_feature      49003..52020
                    /note="assembly_fragment"
misc_feature      52121..57368
                    /note="assembly_fragment"
misc_feature      57469..65330
                    /note="assembly_fragment"
misc_feature      65431..74541
                    /note="assembly_fragment"
misc_feature      74642..85758
                    /note="assembly_fragment"
misc_feature      85859..96006
                    /note="assembly_fragment"

Query Match      7.8%; Score 45.4; DB 2: Length 183345;
Best Local Similarity 49.4%; Pred. No. 0.33;
Matches 118; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

OY 332 AAAACGATAAATCGTGGGACAAATAATAGTCTACCAACCCCATAGAGTAGAGAA 391
    ||| ||||| ||| ||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6437 AATAAATAAATTTCAAAAAGAAATATATATATATATATATATATATATATATAT 6496
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 392 ATATGAACATGATTTAAACAAGAGAGAGTAGCTACAAATGTCCCTAGATATGACATATA 451
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6497 ATATATATATATATATATATACAGCAAGATGGGTGTTGGTAGGGCAATAGAGAAAGTG 6556
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 452 TATCAAGAAACAGGTACTATGATCAATAAATACTAACAGCAATGGGACGAGTGATATAAAAC 511
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6557 TGTATATCAAGAGTAGTGACTTTAGAAATAAATACTAACACAAATAAAAGTAAAAATCTC 6616
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 512 ACTATATGCTGTGGCTCGCTGTTGTTAAATTTGCTTAATAGTATTTTCTCTGGAATTCAT 570
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6617 ACTTAACCATGACTGCTCATTTAAGACTAGAGTAATTCATGTATCTTGTATTTTAT 6675
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 8
AC107857
LOCUS      AC107857      203422 bp      DNA      linear      HTG 26-APR-2002
DEFINITION Mus musculus clone RP23-18K24, WORKING DRAFT SEQUENCE, 12 ordered
            pieces.
ACCESSION AC107857
VERSION   AC107857.3 GI:20330954
KEYWORDS  HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE    Mus musculus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 203422)
AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE     Mus musculus, clone RP23-18K24
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 203422)

```

AUTHORS

Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collimore,A., Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyne,S., Ginde,S., Gird,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G., MacClean,C., MacDonald,P., Major,J., Marquis,N., Matthews,L., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Petrosian,K., Phunkhang,P., Pierre,N., Pollara,V., Ramon,J., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE

Direct Submission

Submitted (24-JAN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 203422)

REFERENCE

Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collimore,A., Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyne,S., Ginde,S., Gird,S., Goyette,M., Graham,L., Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J., Levine,R., Liu,G., MacClean,C., MacDonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

Submitted (26-APR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 26, 2002 this sequence version replaced gi:20258511.
All repeats were identified using RepeatMasker:

COMMENT

Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information

Center project name: L20637

Center clone name: 18_K24

----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 197973 bases at least Q40
Consensus quality: 200612 bases at least Q30
Consensus quality: 201671 bases at least Q20
Insert size: 199000; agarose-fp
Insert size: 202322; sum-of-contigs
Quality coverage: 6.3 in Q20 bases; agarose-fp

Query Match	7.88	Score 45.4	DB 2
Best Local Similarity	49.48	Pred. No. 0.33	

CDS
join(34837..35014,35188..35398,35512..36769)
/note="ORF_id:dd_00653"
/codon_start=1
/product="hypothetical protein"
/protein_id="AAL99332.1"
/db_xref="GI:19807763"
/translation="MSEIDNTDIKKEMDNKVRDPTDNKELLKWTKKKKRNSNDSHI
NFIYRPNFKLLANWKNKATKELTRVLLDHDGFLRIELPDNYLCPTLIRINLY
WISQDLNKLILNDNDNKKIIGIDIGTSCIPFLGAKLFNWSFGIDIDDDV
LEYAONNITINSNLKILFKNEKSDILLKILNYKEGNTFNSNDHDDHDDDD
DEYFADFCICNPPFPKDLNENNNNNKPKSCTGVNEMVTGCGFEFVKRIKES
FOLKCKIRPYTTHMIGKYNLPLINILIKQYILPKNQIOTTELVOGTSRWLWYFL
NKSLETENNN
IVGSSIKLDRDIEFLTFIDLTILFLFKPDKPKINNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNSDCGEIINSNLFILFKLEIKNEIKLK"
37054..38214
/note="ORF_id:dd_01574"
/codon_start=1
/product="PAP-1. 6/101"
/protein_id="AAL99333.1"
/db_xref="GI:19807764"

CDS
/translation="MDRDYSEKVKYKVDNDKDRFSNDKIRHNFDEFTMPNSDRDR
DKDIDYRDYRDYRDYRDYRDYRDYRDYRDYRDYRDYRDYRDYRDYRDYRDY
NNSNSSSFDGDKYRDYRDYRDYRDYRDYRDYRDYRDYRDYRDYRDYRDYRDY
GRSKNKTOYLKEMVANSNGKRELEDSIPDLPENREAREFLRNAPYGLHPLGK
EVKVMOCWCKEFGHRAIECTQTNPIAEYELKREDPMQIYNDHDDFGFNKNKNNIN
NNNNNNNFIHDKRINKINTNDENNNDNNNNNNNNNNNNNNNNNNNNNNNNNN
EKRRKREKKIKASKSHRHSDDSDSGGSSDSDSLKDESYSSNESE"
join(39790..40108,40252..47744)
/note="ORF_id:dd_01572"
/codon_start=1
/product="Putative Acyltransferase"
/protein_id="AAL99334.1"

Query Match 7.7% Score 44.8; DB 2; Length 105470;
Best Local Similarity 51.5% Pred. No. 0.47; Indels 0; Gaps 0;
Matches 103; Conservative 0; Mismatches 97

Qy 376 CATAGAAGTAGAAAAATGAACACTGATTTAAACAAAGAGAGAGTAGCTACAAATGTGCC 435
Db 4177 CACAAATATAGATATTTTCGTACACACTCATCCACACACAAAAATATCTCTGATTA 4118
Qy 436 TAGATATGACAAATATATCAAGAAACAGAGTACTGCATAATAACTAACCAAGATGG 495
Db 4117 TAATATATGTAATAATAATAACAAAAACGATATGTTAGAAAATAAAAAACATATTC 4058
Qy 496 GACGAGTGATAAAACACTATATCTGTGCTCGCTTGAATTTGCTTAATAGTATTT 555
Db 4057 ATCCAAATTTCCACCAATATGTTTTGATTTTCCACTACAAAAAATTTATATTCATTT 3998
Qy 556 TGTCTTGATTTTCAATTTAC 575
Db 3997 TTTTTTTTTTTTTTTTAAAC 3978

RESULT 11
AX281294/c
LOCUS AX281294 6641 bp DNA linear PAT 02-NOV-2001
DEFINITION Sequence 36 from Patent WO0171164.
ACCESSION AX281294
VERSION AX281294.1 GI:16608549
KEYWORDS

SOURCE synthetic construct.
ORGANISM synthetic construct
REFERENCE 1
AUTHORS Olek.A., Piepenbrock,C. and Berlin,K.
TITLE Diagnosis of diseases associated with apoptosis
JOURNAL Patent: WO 0171164-A 36 18-OCT-2001;
Epigenomics AG (DE)
FEATURES Location/Qualifiers
source
1..6641
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT 2495 a 22 c 984 g 3140 t
ORIGIN

Query Match 7.6% Score 44.4; DB 6; Length 6641;
Best Local Similarity 48.8% Pred. No. 0.6;
Matches 120; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

Qy 333 AAACGATATAAATCGGTGGGACAAATAATAGTCTACCAACGCCATAAGAGTAGAAAAA 392
Db 319 AAATATAACAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 260
Qy 393 TATGAACACTGATTTAAACAAAGAGAGTAGCTACAAATTTGTCCTAGATATGACAATAT 452
Db 259 TTCCCAATACCAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 200
Qy 453 ATCAAGAAACAGGTACTATGCATATAAATAACTAACGACGATGGAGAGTGATAAAAAA 512
Db 199 TACTAAACACAATCTAATAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 140
Qy 513 CTATATGCTGTCGTCGCTTGAATTTGCTTAATAGTATTTGCTTCTGATTTTCATTT 572
Db 139 ATATATATTTTCGTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 80
Qy 573 TACCAA 578
Db 79 TACAAA 74

RESULT 12
AX345217/c
LOCUS AX345217 6641 bp DNA linear PAT 01-FEB-2002
DEFINITION Sequence 288 from Patent WO0200928.
ACCESSION AX345217
VERSION AX345217.1 GI:18493103
KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct
REFERENCE 1
AUTHORS Olek.A., Piepenbrock,C. and Berlin,K.
TITLE Diagnosis of diseases associated with the immune system
JOURNAL Patent: WO 0200928-A 288 03-JAN-2002;
Epigenomics AG (DE)
FEATURES Location/Qualifiers
source
1..6641
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"
BASE COUNT 2495 a 22 c 984 g 3140 t
ORIGIN

Query Match 7.6% Score 44.4; DB 6; Length 6641;
Best Local Similarity 48.8% Pred. No. 0.6;
Matches 120; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

Qy 333 AAACGATATAAATCGGTGGGACAAATAATAGTCTACCAACGCCATAAGAGTAGAAAAA 392
Db 319 AAATATAACAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 260
Qy 393 TATGAACACTGATTTAAACAAAGAGAGTAGCTACAAATTTGTCCTAGATATGACAATAT 452

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 24, 2002, 19:31:45 ; Search time 136.703 seconds
(without alignments)
9587.638 Million cell updates/sec

Title: US-10-068-080-12
Perfect score: 582
Sequence: 1 tacaataacgaataacta.....gatttcattttaccaatatt 582

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*
2:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*
3:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*
4:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*
5:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*
6:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*
7:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:*
8:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*
9:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*
10:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:*
11:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*
12:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*
13:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*
14:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*
15:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*
16:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*
17:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*
18:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*
19:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*
20:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*
21:	/SID22/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*
22:	/SID22/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:*
23:	/SID22/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:*
24:	/SID22/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
C 1	46.6	8.0	12025	ABL33298	Human immune syste
C 2	44.4	7.6	6641	ABL54336	Chemically treated
C 3	44.4	7.6	6641	ABL32315	Human immune syste
C 4	40.4	6.9	739	ABO65485	Arabidopsis thalia
C 5	40	6.9	7511	ABL33283	Human immune syste
C 6	39.8	6.8	18855	ABL32611	Human immune syste
C 7	39.6	6.8	542	ABQ28060	Oligonucleotide fo
C 8	39.6	6.8	542	ABQ28061	Oligonucleotide fo
C 9	39.6	6.8	5442	ABL33968	Human immune syste

C 10	39.4	6.8	6802	24	ABL33231	Human immune syste
C 11	39.2	6.7	6130	22	AAS46645	Tumour suppressor
C 12	39.2	6.7	6130	24	ABL32128	Human immune syste
C 13	39	6.7	5379	24	ABL70370	Chemically treated
C 14	39	6.7	5379	24	ABL33677	Human immune syste
C 15	39	6.7	5379	24	ABL34577	Human immune syste
C 16	38.8	6.7	840	21	ABN81104	Shrimp polynucleot
C 17	38.6	6.6	1182	22	ABA82969	Enterococcus faeca
C 18	38.6	6.6	4234	20	AA13250	Enterococcus faeca
C 19	38.4	6.6	1201	24	ABQ14374	Oligonucleotide fo
C 20	38.4	6.6	1201	24	ABQ14375	Oligonucleotide fo
C 21	38.4	6.6	2596	24	ABN59776	Novel human coding
C 22	38.4	6.6	7667	22	AAS46334	Tumour suppressor
C 23	38.4	6.6	18357	24	ABO67083	Human angioogenesis
C 24	38	6.5	788	22	AAH06825	Human CDNA clone (
C 25	38	6.5	2026	22	AAH17925	Human CDNA sequenc
C 26	38	6.5	6243	20	AAZ09598	Clostridium specie
C 27	38	6.5	11944	24	ABL34186	Human immune syste
C 28	37.8	6.5	12356	22	AAS46510	Tumour suppressor
C 29	37.8	6.5	14920	24	ABN80146	Human chemically m
C 30	37.6	6.5	11996	24	ABL34492	Human metastasis a
C 31	37.6	6.5	83391	24	ABO67094	Human angioogenesis
C 32	37.4	6.4	5153	24	ABL32955	Human immune syste
C 33	37.4	6.4	6775	24	ABO67159	Human angioogenesis
C 34	37.4	6.4	1664976	19	AAV21209	Methanococcus jann
C 35	37.2	6.4	898	12	AAQ11132	Skin corn formatio
C 36	37.2	6.4	5647	24	ABL70355	Chemically treated
C 37	37.2	6.4	5647	24	ABL33566	Human immune syste
C 38	37.2	6.4	5647	24	AA561320	Human gene regulat
C 39	37.2	6.4	6306	22	AAS45516	Chemically pretrea
C 40	37.2	6.4	6306	24	ABK28458	DNA transcription
C 41	37.2	6.4	6668	24	ABL33697	Human immune syste
C 42	37.2	6.4	19087	24	ABL32793	Human immune syste
C 43	37.2	6.4	38342	22	AAS46745	Tumour suppressor
C 44	37.2	6.4	38342	24	ABK31506	Signal transductio
C 45	37	6.4	5695	22	AAS46537	Tumour suppressor

ALIGNMENTS

RESULT 1
ABL33298/C
ID ABL33298 standard; DNA; 12025 BP.
AC ABL33298;
XX
XX 26-MAR-2002 (first entry)
DT Human immune system associated gene SEQ ID NO: 1271.
DE
DE Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiatherosclerotic; antianemic; cytostatic; neurotropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antineurotic; antiarthritic; antidiabetic; antipsoriatic;
KW antineurotic; antiarthritic; antidiabetic; antipsoriatic;
KW acute myeloid leukaemia; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;...;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.
XX
XX Homo sapiens.
XX
XX WO200200528-A2.
XX
XX 03-JAN-2002.
XX
XX 02-JUL-2001; 2001WO-EP07537.
XX
XX 30-JUN-2000; 2000DE-1032529.
XX
XX 01-SEP-2000; 2000DE-1043626.
XX
XX (EPIG-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K;
 XX WPI; 2002-130909/17.
 XX
 XX Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation
 XX
 XX Claim 1; SEQ ID NO 1271; 32pp + Sequence Listing; German.
 XX
 XX The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.
 XX
 XX Sequence 12025 BP; 3767 A; 84 C; 2123 G; 6051 T; 0 other;
 SQ
 Query Match 8.0%; Score 46.6; DB 24; Length 12025;
 Best Local Similarity 58.2%; Pred. No. 0.0054;
 Matches 82; Conservative 0; Mismatches 59; Indels 0; Gaps 0;
 QY 380 AGAAGTAGAAAAATATGAACCTGATTTAAACAAAGAGAGAGAGTACCAATTTGCCCTAGA 439
 DB 1407 ATAATATAAAAAATATAATTTATTTAAATATAAAAAACCAATTTATTTTAAATAAT 1348
 QY 440 TATGCAATATATATCAAGAAACAGGTACTATGCAATAATACTACAGCAATGGGACG 499
 DB 1347 TACCAAAATATAAAATATAAATATTAAAAAAATCAAAAAATATAAAAAATAAAAATA 1288
 QY 500 AGTGATAAAAAACACTATATGT 520
 DB 1287 AATAATACAAAACAAATATTT 1267
 RESULT 2
 ABL54336/c
 ID ABL54336 standard; DNA; 6641 BP.
 AC ABL54336;
 XX
 XX 29-JUL-2002 (first entry)
 DT
 DE Chemically treated apoptosis gene complementary to gene #18.
 XX
 XX Apoptosis; HIV; Bloom syndrome; cardiopathy;
 KW neurodegenerative disorder; Herpes simplex virus; renal ischaemia;
 KW amyotrophic lateral sclerosis; cancer; ds.
 XX
 XX Unidentified.
 OS
 PN WO200177164-A2.
 PD
 PD 18-OCT-2001.
 XX
 XX 06-APR-2001; 2001WO-EP03969.
 XX
 XX 06-APR-2000; 2000DE-1019058.
 PR 07-APR-2000; 2000DE-1019173.
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 XX (EPIG-) EPIGENOMICS AG.
 PA
 XX Olek A, Piepenbrock C, Berlin K;
 XX WPI; 2002-017444/02.
 XX
 XX Chemically modified sequences of genes associated with apoptosis are
 PT useful to determine methylation patterns of genomic DNA samples for

PT diagnosis of associated diseases such as cancer
 XX
 XX Claim 1; Seq ID #36; 24pp; English.
 XX
 XX This invention relates to chemically pre-treated DNA of genes
 CC associated with apoptosis. The nucleic acids are used to allocate
 CC patients for specific therapy for HIV infection, Bloom syndrome,
 CC cardiopathy, aging, neurodegenerative disorders, Herpes simplex virus
 CC infection, renal ischaemia, amyotrophic lateral sclerosis, solid tumours
 CC and cancers. This nucleotide sequence represents a chemically
 CC treated apoptosis gene. Even SEQ ID numbers are the complementary
 CC DNA strands to the odd SEQ ID numbers. The sequence data for this
 CC patent is not represented in the printed specification but is based on
 CC information supplied by the European patent office.
 XX
 XX Sequence 6641 BP; 2495 A; 22 C; 984 G; 3140 T; 0 other;
 SQ
 Query Match 7.6%; Score 44.4; DB 24; Length 6641;
 Best Local Similarity 48.8%; Pred. No. 0.018;
 Matches 120; Conservative 0; Mismatches 126; Indels 0; Gaps 0;
 QY 333 AAACGATAAAATGCGTGGGACAAATAAATAGTCTACCAACGCCATAGAAGTAGAAAAA 392
 DB 319 AACTATATAACAAAAAATAAATAAATAAACAACAAATACCAAAAAATTAACATTA 260
 QY 393 TATGAACCTGATTTAAACAAAGAGAGAGTACGTACAAATTTGCCCTAGATATGACAATATAT 452
 DB 259 TTCCCAATACCACAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 200
 QY 453 ATCAAGAAACAGGTACTATGCTATAAATACTACACCAATGGGACGAGTATAAACA 512
 DB 199 TACTAAACACAACTCTAATATATAAATAAATAAATAAATAAATAAATAAATAAATAA 140
 QY 513 CTATATCTGTGCTCGCTTGTAAATTTGCTTAATAGTTTGTCTTGGATTTCATTT 572
 DB 139 ATATATATTATTCGCTAAACACATAATAAATAAATAAATAAATAAATAAATAAATAA 80
 QY 573 TACCAA 578
 DB 79 TACAAA 74
 RESULT 3
 ABL32315/c
 ID ABL32315 standard; DNA; 6641 BP.
 AC ABL32315;
 XX
 XX 26-MAR-2002 (first entry)
 DT
 DE Human immune system associated gene SEQ ID NO: 288.
 XX
 XX Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antianaemic; cytostatic; neurotropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.
 XX
 XX Homo sapiens.
 OS
 XX WO200200928-A2.
 PN
 XX 03-JAN-2002.
 PD
 XX 02-JUL-2001; 2001WO-EP07537.
 PF
 XX 30-JUN-2000; 2000DE-1032529.
 PR
 XX 01-SEP-2000; 2000DE-1043826.
 XX
 XX (EPIG-) EPIGENOMICS AG.
 PA

XX Olex A, Piepenbrock C., Berlin K;
 XX WPI: 2002-130909/17.
 XX Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation
 XX Claim 1; SEQ ID NO 288; 32pp + Sequence Listing; German.
 XX The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.
 XX
 SQ Sequence 6641 BP; 2495 A; 22 C; 984 G; 3140 T; 0 other;
 Query Match 7.6%; Score 44.4; DB 24; Length 6641;
 Best Local Similarity 48.8%; Pred. No. 0.018;
 Matches 120; Conservative 0; Mismatches 126; Indels 0; Gaps 0;
 Qy 333 AAACGATATAAATGCTGGGACATAAATAGTGTACCAAGCCCATAGAGTAGAAAAA 392
 Db 319 AAATATATAACAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 260
 Qy 393 TATGAATGATTTAAACAAAGAGAGAGTAGTACAAATTTGCCCTAGATATGACAAATAT 452
 Db 259 TTCCCAATACCAAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 200
 Qy 453 ATCAAGAAACAGTACTATGTCATAAATACTACACCAATGGGACGAGTGATAAAAACA 512
 Db 199 TACTAAACACAACTCTAATAAATTTAATAAATAAATAAATAAATAAATAAATAAATA 140
 Qy 513 CTATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 572
 Db 139 ATATATATATTCCTCAAAACATATAAATAAATAAATAAATAAATAAATAAATAAATA 80
 Qy 573 TACCAA 578
 Db 79 TACAAA 74
 RESULT 4
 ABQ65485
 XX ABQ65485 standard; DNA; 739 BP.
 XX AC ABQ65485;
 XX DT 21-AUG-2002 (first entry)
 XX Arabidopsis thaliana polynucleotide SEQ ID NO 62.
 DE Arabidopsis thaliana; thale cress; plant; transgenic; GWO; disease;
 KW stress; metabolic pathway; biosynthetic pathway; nutrition; fungicide;
 KW insecticide; antibiotic; ds.
 XX OS Arabidopsis thaliana.
 XX PN US2002059663-A1.
 XX PD 16-MAY-2002.
 XX PF 26-JAN-2001; 2001US-0770149.
 XX PR 27-JAN-2000; 2000US-178506P.
 XX (GORLACH J.
 PA (ANY)/ AN Y.

PA (HAMI/) HAMILTON C M.
 PA (PRIC/) PRICE J L.
 PA (RAIN/) RAINES T M.
 PA (YUYX/) YU Y.
 PA (RAME/) RAMEAKA J G.
 PA (PAGE/) PAGE A.
 PA (MATH/) MATHAW A V.
 PA (LEDF/) LEDFORD B L.
 PA (WOES/) WOESSNER J P.
 PA (HAAS/) HAAS W D.
 PA (GARC/) GARCIA C A.
 PA (KRICK/) KRICKER M.
 PA (SLAT/) SLATER T.
 PA (DAVI/) DAVIS K R.
 PA (ALLE/) ALLEN K.
 PA (HOPE/) HOFFMAN N.
 PA (HURB/) HURBAN P.
 XX
 PI Gorlach J, An Y, Hamilton CM, Price JL, Raines TM, Yu Y;
 PI Rameaka JG, Page A, Mathew AV, Ledford BL, Woessner JP, Haas WD;
 PI Garcia CA, Kricker M, Slater T, Davis KR, Allen K, Hoffman N;
 PI Hurban P;
 XX WPI: 2002-479224/51.
 DR New nucleic acid that hybridizes to Arabidopsis thaliana sequences,
 XX useful e.g. for preparing transgenic plants with increased resistance
 PT or altered metabolism
 PT
 XX Claim 1; SEQ ID NO 62; 40pp + Sequence Listing; English.
 PS The invention relates to nucleic acids (I) that hybridise under stringent
 CC conditions to any of 999 sequences (ABQ65424-ABQ66422) or their
 CC fragments. (I) are used to express the corresponding polypeptides (II) or
 CC to produce genetically modified plant cells or transgenic plants, which
 CC may have improved resistance to disease or stress, or altered
 CC metabolic/biosynthetic pathways (for production of commercial,
 CC nutritional or medicinal products), or generally any trait of interest,
 CC or can be used to screen for biologically active agents (e.g. fungicides,
 CC insecticides and antibiotics)
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC USPTO at seqdata.uspto.gov/sequence.html?DocID=999909770149.
 XX
 SQ Sequence 739 BP; 289 A; 163 C; 134 G; 146 T; 7 other;
 Query Match 6.9%; Score 40.4; DB 24; Length 739;
 Best Local Similarity 53.9%; Pred. No. 0.11;
 Matches 83; Conservative 0; Mismatches 71; Indels 0; Gaps 0;
 Qy 364 TGCTACCAAGCCCATAGAGTAGTAAAAATATCAACTGATTTAAACAAAGAGAGTAGC 423
 Db 286 TGCATATAATTCAGCAAACTTGGGTAAAGATAAATTTAATTAATGAAGAGTCCATAAC 345
 Qy 424 TACAATGTCCTAGATATGACAATATATATCAAGAAACAGCTACTATGCATAAATAAC 483
 Db 346 TGATGCTCTACCATATTGAAACTCTAAAAACCATAGACCAGCAGCAAGAAAGCTATTA 405
 Qy 484 TACACCAATGGGACGAGTGATATAAACACTATA 517
 Db 406 CATCAGCATAGGAAAGAGCTATATAATCACCATA 439
 RESULT 5
 ABL33283/c
 ID ABL33283 standard; DNA; 7511 BP.
 XX
 XX ABL33283;
 XX AC
 XX DT 26-MAR-2002 (first entry)
 XX Human immune system associated gene SEQ ID NO: 1256.
 XX

KW Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antianemic; cytosine; cytosine; cytosine;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.
 XX Homo sapiens.
 OS
 XX WO20020928-A2.
 PN
 XX 03-JAN-2002.
 PD
 XX 02-JUL-2001; 2001WO-EP07537.
 XX
 XX 30-JUN-2000; 2000DE-1032529.
 PR
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 XX (EPIG-) EPIGENOMICS AG.
 PA
 XX Olek A, Piepenbrock C, Berlin K;
 PI
 XX WPI; 2002-130909/17.
 DR
 XX Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation
 PT
 XX Claim 1; SEQ ID NO 1256; 32pp + Sequence Listing; German.
 PS
 XX The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.
 XX
 XX Sequence 7511 BP; 2046 A; 93 C; 1664 G; 3708 T; 0 other;
 SQ

Query Match 6.9%; Score 40; DB 24; Length 7511;
 Best Local Similarity 51.7%; Pred. No. 0.37;
 Matches 91; Conservative 0; Mismatches 85; Indels 0; Gaps 0;
 OY 406 AACAAAGAGAGAGTACAAATGTCCTAGATATGACAAATATATATCAAGAAACAG 465
 Db 6546 AACCTAAAGGACAAACGAACTCCGTCTCAAAAAAATAAAAAAATAAAAAA 6487
 OY 466 GTACTATGCATAATAACTACACCAATGGGACGAGTGATAAACACTATATCTGTG 525
 Db 6486 AACCCACATATATATTAATTAACAAACCAAAAAAATAAAAAAATAAAAAA 6427
 OY 526 GCTCGCTGAATGCTATATGCTTATGCTTGGATTTTCAATTTACCAATAT 581
 Db 6426 TTCTACTTTTATTTCTATTTATTTATTTTATTAATTAATTTATTAACATAAT 6371

RESULT 6
 ABL32611/c
 ID ABL32611 standard; DNA; 18855 BP.
 XX ABL32611;
 AC
 XX 26-MAR-2002 (first entry)
 DT
 XX Human immune system associated gene SEQ ID NO: 584.
 DE
 XX Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antianemic; cytosine; cytosine; cytosine;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;

KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.
 XX Homo sapiens.
 OS
 XX WO20020928-A2.
 PN
 XX 03-JAN-2002.
 PD
 XX 02-JUL-2001; 2001WO-EP07537.
 XX
 XX 30-JUN-2000; 2000DE-1032529.
 PR
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 XX (EPIG-) EPIGENOMICS AG.
 PA
 XX Olek A, Piepenbrock C, Berlin K;
 PI
 XX WPI; 2002-130909/17.
 DR
 XX Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation
 PT
 XX Claim 1; SEQ ID NO 584; 32pp + Sequence Listing; German.
 PS
 XX The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.
 XX
 XX Sequence 18855 BP; 5332 A; 178 C; 4053 G; 9292 T; 0 other;
 SQ

Query Match 6.8%; Score 39.8; DB 24; Length 18855;
 Best Local Similarity 48.9%; Pred. No. 0.61;
 Matches 107; Conservative 0; Mismatches 112; Indels 0; Gaps 0;
 OY 355 AATAAATAGTGTACCAACGCCATAGAGTACAAAATATGAACTGATTTAAACAAAGA 414
 Db 11201 ACTAATTCATTAACCTTTACACAAACCCCTAAATAAATAAATCTCCCTCCACATATA 11142
 OY 415 GAGAGTAGCTACAAATTTGTCCTAGATATGACAATATATATCAAGAAACAGGTACTATGC 474
 Db 11141 AACAAATCAATTAATATCAACAAATATAAACAACAAAAAACAACCGTCAACATAT 11082
 OY 475 ATAAATAACTAACGCAATGGGACGAGTGATATAAACACTATATGCTGTGGCTCGCTTG 534
 Db 11081 TATACTAATTTTTTTTAAAAATAAAAAATATAAAAAAATAAATACTATTTTTTATTTTC 11022
 OY 535 TAATTTGCTTAATAGTATTTTCTGTTGGATTTTCATTTT 573
 Db 11021 TAAATTTCCAAATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 10983

RESULT 7
 ABQ28060/c
 ID ABQ28060 standard; DNA; 542 BP.
 XX ABQ28060;
 AC
 XX 12-JUL-2002 (first entry)
 DT
 XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 14651.
 DE
 XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
 KW drug; side effect; cancer; central nervous system; cardiovascular;
 KW

Db 6273 ATCTAAAAAACTTAAACAAACAAATAAATACTTCAATACATAAAAAAATAAAAAAAT 6214
 QY 394 ATGAACGATTTAAACAAAGAGAGAGTAGTACAAATGTCCCTAGATATGACAAATATA 453
 Db 6213 AATAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 6154
 QY 454 TCAAGAAACAGGTACTTATCATATAAATAAATAACAGCAATGGGACGAGTATAAAAA 510
 Db 6153 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 6097
 RESULT 11
 AAS46645/c
 ID AAS46645 standard; DNA; 6130 BP.
 AC AAS46645;
 XX
 DT 18-DEC-2001 (first entry)
 DE Tumour suppressor gene derived chemically modified sequence #367.
 XX
 KW Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
 KW cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
 KW cytosine methylation; ds.
 OS Homo sapiens.
 XX
 PN W0200168912-A2.
 XX
 PD 20-SEP-2001.
 XX
 PF 15-MAR-2001; 2001WO-EP02955.
 XX
 PR 15-MAR-2000; 2000DE-1013847.
 PR 06-APR-2000; 2000DE-1019058.
 PR 07-APR-2000; 2000DE-1019173.
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 PA (EPIG-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 DR WPI; 2001-602752/68.
 XX
 PT Fragments of chemically modified genes associated with tumour suppressor
 PT genes and oncogenes, useful in designing primers and probes for
 PT analysing diseases associated with cytosine methylation state e.g.
 PT cancer
 XX
 PS Claim 1; SEQ ID NO 367; 27pp; English.
 XX
 CC The invention relates to a nucleic acid comprising a sequence of 18
 CC bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
 CC bisulphite, of genes associated with tumour suppression and
 CC oncogenes having a sequence taken from 536 (actually 533 since
 CC numbers 408, 458 and 500 are missing from the sequence listing) sequences
 CC (Ss) and sequences complementary to (Ss). The nucleic acid may be a
 CC peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
 CC form part of a set of probes for detecting the cytosine methylation state
 CC and/or single nucleotide polymorphisms and also to be used in an
 CC array for analysing diseases associated with CpG dinucleotides e.g.
 CC cancers and tumours. The probes can also be used in a method for
 CC ascertaining genetic and/or epigenetic parameters for the diagnosis
 CC and/or therapy of existing diseases or the predisposition to specific
 CC diseases, by analysing cytosine methylations. The parameters may be
 CC compared to another set of genetic and/or epigenetic parameters, the
 CC differences serving as basis for diagnosis and/or prognosis events which
 CC are disadvantageous to patients. The present sequence is one of the
 CC 533 genomic sequences derived from tumour suppressor genes and
 CC oncogenes.
 CC Note: The sequence data for this patent did not form part

CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 6130 BP; 1489 A; 244 C; 1571 G; 2826 T; 0 other;
 Query Match 6.7%; Score 39.2; DB 22; Length 6130;
 Best Local Similarity 54.1%; Pred. No. 0.58;
 Matches 80; Conservative 0; Mismatches 68; Indels 0; Gaps 0;
 QY 433 CCTAGATATGACAATATATATCAAGAAACAGGTACTATGTCATATAATACTAACACAA 492
 Db 650 CTCAATATAACAAATAAAAAATAAAACACACAAACAAATAATACCGTACATCCA 591
 QY 493 TGGGACGAGTGATAAAAAACACTATATGCTGTGCTCGCTTGTAAATTTGCTTAATAGTTA 552
 Db 590 AAGCATAAAAACTAAAAAACCATATTTCTATATTTATTACTATTATTATTTAATAAA 531
 QY 553 TTTTGTCTTGGATTTCATTTTACCATA 580
 Db 530 CTTTATTTTATTTTATTTTATTTTAAAAA 503
 RESULT 12
 ABL32128/c
 ID ABL32128 standard; DNA; 6130 BP.
 AC ABL32128;
 XX
 DT 26-MAR-2002 (first entry)
 DE Human immune system associated gene SEQ ID NO: 101.
 XX
 KW Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antianaemic; cytostatic; neutropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.
 XX
 OS Homo sapiens.
 XX
 PN W0200200528-A2.
 XX
 PD 03-JAN-2002.
 XX
 PF 02-JUL-2001; 2001WO-EP07537.
 XX
 PR 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX
 PA (EPIG-) EPIGENOMICS AG.
 XX
 PI Olek A, Piepenbrock C, Berlin K;
 XX
 DR WPI; 2002-130909/17.
 XX
 PT Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation
 XX
 PS Claim 1; SEQ ID NO 101; 32pp + Sequence Listing; German.
 XX
 CC The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.

The present invention provides a number of human immune system associated genes which are modified by the methylation of cytosines. The sequences can be used in the diagnosis and treatment of immune system disorders, including eye diseases such as retinopathy, neovascular glaucoma and macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis, rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel diseases. The present sequence is a gene of the invention.

XX SQ Sequence 5379 BP; 1692 A; 52 C; 1038 G; 2597 T; 0 other;
Query Match 6.7%; Score 39; DB 24; Length 5379;
Best Local Similarity 48.4%; Pred. No. 0.63;
Matches 108; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

QY 360 ATAGTGTCTACCAACGCCATAGAAAGTAGAAAAATATGAACACTGATTTTAAACAAGAGAGAG 419
DB 2554 ATATAAATAATTACTCTTCTAAATTTATTAAAAACAACAACACCATTTAAACTTAATAAAT 2495

QY 420 TAGCTACAATTTGTCCTAGATATGACAATATATCAAAAGAACAGGTACTATGCATATA 479
DB 2494 AAAATAAAAAATTTAAAAATTAATATCAATATCTAAATCATATAAAAAAATAATAAATAAAA 2435

QY 480 TAACTACACCAATGGGACGAGTGATAAAAAACACATATATCTCTGGCTCGCTTGTAAAT 539
DB 2434 AAAATAAATTCATATACTTTTAAATAAAAAATTAATTAATTCGTTAAATTAATTTT 2375

QY 540 TGCTTAATAGTATTTTGTCTTGGATTTCATTTTACCAATATT 582
DB 2374 TTCTAATCAATTATCTTTTAAAAATTTTCTCTCTCTTTCTT 2332

Search completed: December 24, 2002, 22:49:57
Job time : 175.703 secs

RESULT 15
ABL34577/c
ID ABL34577 standard; DNA; 5379 BP.
XX AC ABL34577;
XX DT 26-MAR-2002 (first entry)
XX DE Human metastasis associated gene SEQ ID NO: 130.
XX KW Metastasis associated gene; cytostatic; gene therapy; cancer;
XX KW cytosine methylation; gene; ds.
XX OS Homo sapiens.
XX PN WO200177376-A2.
XX PD 18-OCT-2001.
XX PF 06-APR-2001; 2001WO-EP03970.
XX PR 06-APR-2000; 2000DE-1019058.
XX PR 07-APR-2000; 2000DE-1019173.
XX PR 30-JUN-2000; 2000DE-1032529.
XX PR 01-SEP-2000; 2000DE-1043826.
XX PA (EPIG-) EPIGENOMICS AG.
XX PI Olek A, Piepenbrock C, Berlin K;
XX DR WPI; 2002-010922/01.
XX PT New nucleic acid derived from chemically treated metastasis genes,
XX PT useful for diagnosis of cancers by analysis of cytosine methylation,
XX PT also for treatment
XX PS Claim 1; SEQ ID NO 130; 23pp + Sequence Listing; English.
XX CC The present invention provides a number of human metastasis associated
XX CC genes which are modified by cytosine methylation. The sequences can be
XX CC used in the diagnosis and treatment of cancer. The present sequence is
XX CC one of the genes of the invention.
XX SQ Sequence 5379 BP; 1692 A; 52 C; 1038 G; 2597 T; 0 other;
Query Match 6.7%; Score 39; DB 24; Length 5379;
Best Local Similarity 48.4%; Pred. No. 0.63;
Matches 108; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 24, 2002, 22:16:24 ; Search time 3199.81 Seconds
(without alignments)
2945.728 Million cell updates/sec

Title: US-10-068-080-12

Perfect score: 582

Sequence: 1 tacaattaagaataaacta.....gatttcattttaccaatatt 582

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST: *
1: em_estba: *
2: em_esthum: *
3: em_estmu: *
4: em_estov: *
5: em_estov: *
6: em_estpl: *
7: em_estro: *
8: em_hic: *
9: gb_estl: *
10: gb_est2: *
11: gb_hic: *
12: gb_est3: *
13: gb_est4: *
14: gb_est5: *
15: em_estfun: *
16: em_estom: *
17: gb_gss: *
18: em_gss_hum: *
19: em_gss_inv: *
20: em_gss_pln: *
21: em_gss_vrt: *
22: em_gss_fun: *
23: em_gss_mam: *
24: em_gss_mus: *
25: em_gss_othr: *
26: em_gss_pro: *
27: em_gss_rod: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	51.4	8.8	1030	CNS02LP3	AL203088 Tetraodon
C 2	48.2	8.3	889	CNS006W4	AL065999 Drosophil
C 3	46.8	8.0	777	CNS025WB	AL182612 Tetraodon
C 4	46.2	7.9	997	CNS0134P	AL102403 Drosophil
C 5	45.4	7.8	1201	CNS01660	AL106338 Drosophil
C 6	45.2	7.8	1125	9 AL547503	AL547503 AL547503

7	44.8	7.7	889	17	CNS006W4
8	44.6	7.7	639	17	CNS0170D
C 9	44.2	7.6	1101	17	CNS00L72
C 10	44	7.6	457	17	CNS009E1
C 11	44	7.6	891	17	AZ545935
12	43.6	7.5	877	17	AZ543123
13	43.6	7.5	894	17	AZ687792
C 14	43.4	7.5	836	17	AZ546944
C 15	43.4	7.5	880	17	BH155564
C 16	43.4	7.5	891	17	AZ539310
C 17	43.4	7.5	901	17	AZ548895
C 18	43.4	7.5	932	17	AZ689505
C 19	43.2	7.4	947	17	BH139082
C 20	43.2	7.4	1125	9	AL547503
21	43	7.4	875	17	CNS06MSO
22	43	7.4	928	17	CNS00DKY
C 23	42.8	7.4	832	17	CNS0110K
C 24	42.8	7.4	854	17	AZ688340
25	42.8	7.4	892	17	AZ540933
26	42.8	7.4	934	17	AZ542242
C 27	42.8	7.4	939	17	BH163549
C 28	42.6	7.3	898	17	AZ667815
C 29	42.4	7.3	417	17	CNS02AH7
30	42.4	7.3	1042	17	CNS00192
31	42.2	7.3	881	17	AZ673723
32	42.2	7.3	952	17	CNS006V8
C 33	42	7.2	804	17	CNS03GUX
C 34	42	7.2	847	17	AZ676214
35	42	7.2	917	17	AZ677473
C 36	42	7.2	951	17	AZ676519
37	42	7.2	1058	17	CNS011HR
38	42	7.2	1101	17	CNS00EVL
C 39	41.8	7.2	345	17	CNS04N9U
C 40	41.8	7.2	647	17	BH154473
41	41.8	7.2	799	17	AZ688611
C 42	41.8	7.2	822	17	AZ667519
C 43	41.8	7.2	828	17	AZ533304
44	41.8	7.2	828	17	CNS02PCX
45	41.8	7.2	830	17	AZ678784

ALIGNMENTS

RESULT 1	CNS02LP3/C	CNS02LP3	1030 bp	DNA	linear	GSS 14-MAY-2000
LOCUS	Tetraodon nigroviridis genome survey sequence T7 end of clone 147L04 of library G from Tetraodon nigroviridis, genomic survey sequence.					
DEFINITION	AL203088					
ACCESSION	AL203088.1	GI:7861433				
VERSION	GSS: genome survey sequence.					
KEYWORDS	Tetraodon nigroviridis.					
SOURCE	Tetraodon nigroviridis					
ORGANISM	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes; Tetraodontidae; Tetraodon.					
REFERENCE	1 (bases 1 to 1030)					
AUTHORS	Roest-Crollius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C., Barnot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F., Saurin, W. and Weissenbach, J.					
TITLE	Human gene number estimate provided by genome wide analysis using Tetraodon nigroviridis DNA sequence					
JOURNAL	Unpublished					
REFERENCE	2 (bases 1 to 1030)					
AUTHORS	Roest-Crollius, H., Jaillon, O., Dasilva, C., Fizames, C., Fisher, C., Bouneau, L., Billault, A., Quetier, F., Saurin, W., Barnot, A. and Weissenbach, J.					
TITLE	Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis					
JOURNAL	Unpublished					


```

Qy 398 ACTGATTAAACAAAGAGAGTAGCTACAAATTTGTCCTAGATATGACAATATATATCAAA 457
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1004 AAAAAAAAAAAWATATAATTCCKAAAWAAWTTTWTAAACAAAAAATATTTTAA 1063

Qy 458 AGAAGACAGTACTATGCAATAACTAACAGCAATGGGAGAGTGATAAAACACTATA 517
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1064 AATWAAARAAATTAATAAATAAATAAATAATTTTWTAAATTAATAATTAATAATWTW 1123

Qy 518 TGCTGTCGCTGGCTGTAATTTGCTTAATAGTATTTTGTCTGGATTTTACCA 577
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1124 AAAAAATAATWATTTTWTWAAAAAATAATTTTATATTAATAATTAATAATWTW 1183

Qy 578 ATAT 581
      ||
Db 1184 TWAT 1187

RESULT 6
AL547503/c 1125 bp mRNA linear EST 16-FEB-2001
LOCUS AL547503 LTI_NFL006.PL2 Homo sapiens cDNA clone CS0D1011YB03 5
DEFINITION prime, mRNA sequence.
ACCESSION AL547503
VERSION AL547503.1 GI:12881624
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
JOURNAL Full-length cDNA libraries and normalization
COMMENT Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source
1..1125
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0D1011YB03"
/clone_lib="LTI_NFL006.PL2"
/tissue_type="placenta"
/note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 415 a 142 c 148 g 345 t 75 others
ORIGIN

Query Match 7.8%; Score 45.2; DB 9; Length 1125;
Best Local Similarity 41.7%; Pred. No. 0.44; Indels 0; Gaps 0;
Matches 101; Conservative 20; Mismatches 121;

Qy 332 AAAACGATAAAATGCGTGGGACAATAAATAGTGTACCAAGCCATAGAGTAGAATA 391
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 986 AAAAAAAWATAATTAATAAATAAATAATTTTWTWTTTATATAAATAAATAAATA 927

Qy 392 ATATGAAGTATTAACAAAGAGAGTAGTACAAATTTGTCCTAGATATGACAATATA 451
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 926 AAAAAAAWAAATTTTAAATAAATAAATAAATAAATAAATAAATAAATAAATA 867

Qy 452 TATCAAGAACAGGCTACTATGCAATAATACTACAGCAATGGGAGAGTGATAAACA 511
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 866 WAAAAAANTANTTATATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 807

```

```

Qy 512 ACTATATGCTGTGCTGGCTTGTAAATTTGCTTAATAGTATTTTGTCTTGGATTTCAAT 571
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 806 AATTTTNNANTATTTTTTTTTTTNNNTATTTTNNNTATTTTNNNTATTTTNNNTATTTT 747

Qy 572 TT 573
      ||
Db 746 TT 745

RESULT 7
CNS006W4 889 bp DNA linear GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #
DEFINITION BACR14N09 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL065999
VERSION AL065999.1 GI:4944967
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 889)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osogawa and
Aaron Mammoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
source
1..889
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACR14N09"
/clone_lib="RPCI-98"
/note="end : TET3"

BASE COUNT 338 a 62 c 30 g 387 t 72 others
ORIGIN

Query Match 7.7%; Score 44.8; DB 17; Length 889;
Best Local Similarity 47.0%; Pred. No. 0.55;
Matches 118; Conservative 2; Mismatches 131; Indels 0; Gaps 0;

Qy 332 AAAACGATAAAATGCGTGGGACAATAAATAGTGTACCAAGCCATAGAGTAGAATA 391
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 354 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 413

Qy 392 ATATGAAGTATTAACAAAGAGAGTAGTACCAATTTGTCCTAGATATGACAATATA 451
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 414 ANAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 473

Qy 452 TATCAAGAACAGGCTACTATGCAATAATACTACAGCAATGGGAGAGTGATAAACA 511
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 474 AAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 533

Qy 512 ACTATATGCTGTGCTGGCTTGTAAATTTGCTTAATAGTATTTTGTCTTGGATTTCAAT 571
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 534 ATTTTNTTTTAAATTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTT 593

```

```

Qy 572 TTACCAATATT 582
   ||| |||
Db 594 TTATTTATTT 604

RESULT 8
CNS017QD 639 bp DNA linear GSS 26-JUL-1999
LOCUS Drosophila melanogaster genome survey sequence SP6 end of BAC
DEFINITION BACN37M13 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL108367
VERSION AL108367.1 GI:5628671
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 639)
Genoscope.
Direct Submission
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelobAC11.

FEATURES
    source          Location/Qualifiers
    1..639
        /organism="Drosophila melanogaster"
        /db_xref="taxon:7227"
        /clone="BACN37M13"
        /clone_lib="DrosBAC"
        /plasmid="pBelobAC11"
        /note="end : SP6"
BASE COUNT 389 a 53 c 46 g 52 t 99 others
ORIGIN
Query Match 7.7% Score 44.6; DB 17; Length 639;
Best Local Similarity 38.08; Pred. No. 0.62; Indels 0; Gaps 0;
Matches 68; Conservative 36; Mismatches 75;

Qy 332 AAAAGCAATAAATGCGTGGGACATAAATAGTGTACCAACGCCATAGAGTAGAGAAA 391
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 220 AWRAAAATAAAWATATCAACWAAATAAARTAAATAAATAAATAAATAAATAAATAA 279
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 392 ATATGAAGTATTTAAACAAGAGAGAGTAGCTACAAATGTCCTAGATATGACATATA 451
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 280 WATTTAAATTTTAAAGAAAAAATAAATGCGGAAWAAWAAWAAWAAWAAWAAATGAA 339
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 452 TATCAAGAACAGCTACTATGCTAATAATACTACACACATGGCAGCTGATAAAAA 510
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 340 ARARWAGAGWAGTGGAACACAAWAAWAAWAAWAAWAAWAAWAAWAAWAAWAAW 398
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 9
CNS00LT2/c 1101 bp DNA linear GSS 14-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC:
DEFINITION BACR48P19 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL078714
VERSION AL078714.1 GI:5102004
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 457)
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :

```

```

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1101)
Genoscope.
Direct Submission
Submitted (11-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammoler in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
    source          Location/Qualifiers
    1..1101
        /organism="Drosophila melanogaster"
        /db_xref="taxon:7227"
        /clone="BACR48P19"
        /clone_lib="RPCI-98"
        /note="end : TET3"
BASE COUNT 469 a 6 c 69 g 151 t 406 others
ORIGIN
Query Match 7.6% Score 44.2; DB 17; Length 1101;
Best Local Similarity 40.38; Pred. No. 0.79; Indels 0; Gaps 0;
Matches 89; Conservative 29; Mismatches 103;

Qy 353 ACATAAATAGTGTACCAACGCCATAGAGTAGAGAAAATATGAAGTATTAACAAA 412
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 656 AAAAAAATWTAWAAWAAAAAATTTAAAAAATAAATAAATAAATAAATAAATAAATA 597
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 413 GAGAGAGTAGCTACAAATGTCCTCCATAGATATGACAAATATATCAAGAACAGGTACTAT 472
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 596 WAAAAAATTTAAAAAATAAATTTTAAATTTAAATTTAAATTTAAATTTAAATTTAA 537
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 473 GCATAAATACTAACAGCAATGGCAGCTGATATAAACAATATATGCTCTGGCTCGCT 532
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 536 TAAWAAWTAATAAAAHWMTTMMHHTTAAWAAWAAWAAWAAWAAWAAWAAWAAW 477
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 533 TGTAAATTTGCTTAATAGTTATTTTGTCTGTGGATTTCATTTT 573
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 476 TTTNTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 436
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 10
CNS009EJ 457 bp DNA linear GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC:
DEFINITION BACR19A04 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL053475
VERSION AL053475.1 GI:4934951
KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 457)
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :

```

BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see <http://www.fruitfly.org> The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mamoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

source

1. .457
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACR19A04"
/clone_lib="RPCI-98"
/note="end : TET3"
Location/Qualifiers

BASE COUNT 181 a 51 c 58 g 62 t 105 others

ORIGIN

Query Match 7.68; Score 44; DB 17; Length 457;

Best Local Similarity 36.68; Pred. No. 0.87; Indels 0; Gaps 0;

Matches 68; Conservative 39; Mismatches 79; Indels 0; Gaps 0;

QY 324 TCCTCCGCAACGATAAATGCGTGGGCAATAAATAGTGTACCAACGCCATAAGAA 383

DB 220 TMMAGAAAAMWTAAAWAAWATGCTAAANAAGTGTGCGGCGMAAAWAAAAA 279

QY 384 GTAGAAAATATGACTGATTAAACAAAGAGAGAGTACGATTCCTCCTAGATG 443

DB 280 AAWAAMAAWAWAGCTGGTGAAMAAWAAWAGAAWAAWMAAAWMAAAATATAGG 339

QY 444 ACAATATATACAAGAACAGGCTACTGATGATTAATCAATCAAGCAGCAGGAGTG 503

DB 340 AAWAAMAAWAAWAAWAAWMAAAWAAWAAWAAWAAWAAWAAWAAWAAWAAW 399

QY 504 ATAAAA 509

DB 400 AAWAA 405

RESULT 11

AZ545935/c

LOCUS

DEFINITION ENTFL49TR Entamoeba histolytica Sheared DNA Entamoeba histolytica

genomic, DNA sequence.

ACCESSION AZ545935

VERSION AZ545935.1 GI:11166985

KEYWORDS GSS.

SOURCE Entamoeba histolytica.

ORGANISM Entamoeba histolytica.

REFERENCE 1 (bases 1 to 891)

AUTHORS Loftus, B., Van Aken, S. and Fraser, C.

TITLE Determination of clone end sequences from Entamoeba histolytica

JOURNAL HMI:IMSS sheared DNA library

COMMENT Unpublished (2000).

Contact: Brendan J Loftus

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0208

Fax: 301 838 3543

Email: bjloftus@tigr.org

Clones are derived from the Entamoeba histolytica HMI:IMSS sheared

DNA library

Seq primer: M13-Reverse
Class: Shotgun
High quality sequence start: 38
High quality sequence stop: 706.
Location/Qualifiers

FEATURES

source

1. .891
/organism="Entamoeba histolytica"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: PHOSI; Site_1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. W. Vaudin and B.
Barell, Oxford University Press, 1999)."

BASE COUNT 277 a 95 c 80 g 439 t

ORIGIN

Query Match 7.68; Score 44; DB 17; Length 891;

Best Local Similarity 51.08; Pred. No. 0.88;

Matches 104; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

QY 371 AACCCCATAGAGACGATAAATATGAACCTGATTAAACAAAGAGAGTAGCTACAATT 430

DB 545 AAAGAAATAGATAAATAGATAAGATGAATGAATGAATGAATGAATGAATGAAT 486

QY 431 GTCCCTAGATAGACATATATATCAAGAAACAGGCTACTGATGATTAATCAATCAAGC 490

DB 485 TCATTGAATTAATTCATTATGATCAATGAATGAATGAATGAATGAATGAATGAAT 426

QY 491 AATGGGACGAGTGAATAAACACATATATGCTGGGCTCGCTGCTTAATTTCTTAATAGT 550

DB 425 AATAAAGAAATGAATAAATATTTATTTATTTATTTATTTATTTATTTATTTATTT 366

QY 551 TATTTCTCTTGGATTTCAATTTA 574

DB 365 ATTCTTATTTTATTTCTTATTTA 342

RESULT 12

AZ543123

LOCUS

DEFINITION

genomic, DNA sequence.

ACCESSION AZ543123

VERSION AZ543123.1 GI:11161750

KEYWORDS GSS.

SOURCE Entamoeba histolytica.

ORGANISM Entamoeba histolytica.

REFERENCE 1 (bases 1 to 877)

AUTHORS Loftus, B., Van Aken, S. and Fraser, C.

TITLE Determination of clone end sequences from Entamoeba histolytica

JOURNAL HMI:IMSS sheared DNA library

COMMENT Unpublished (2000).

Contact: Brendan J Loftus

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0208

Fax: 301 838 3543

Email: bjloftus@tigr.org

Clones are derived from the Entamoeba histolytica HMI:IMSS sheared

DNA library

Seq primer: M13-Forward

Search completed: December 25, 2002, 14:17:29
Job time : 3205.81 secs

Db 361 TAGTGCTACCAAGCCATAGAAAGTATGAAGTATTAAACAAAGAGAGT 420
Qy 421 AGCTACAATTCCTCTAGATGACATATATATCAAGAAACAGGTACTATGATTAAT 480
Db 421 AGCTACAATTCCTCTAGATGACATATATATCAAGAAACAGGTACTATGATTAAT 480
Qy 481 AACTAAACAGCAATGGGAGGAGTATATAAACAACATATATGCTGTGGCTGCTGTAATTT 540
Db 481 AACTAAACAGCAATGGGAGGAGTATATAAACAACATATATGCTGTGGCTGCTGTAATTT 540
Qy 541 GCTTAATAGTATTTTGTCTGGATTTCAATTTTACCAATAT 582
Db 541 GCTTAATAGTATTTTGTCTGGATTTCAATTTTACCAATAT 582

RESULT 2

US-09-222-938A-15

; Sequence 15, Application US/09222938A

; Patent No. 6437108

; GENERAL INFORMATION:

; APPLICANT: Youngman, Philip

; APPLICANT: Fritz, Christian

; APPLICANT: Murphy, Christopher

; APPLICANT: Guzman, Luz-Maria

; TITLE OF INVENTION: ESSENTIAL BACTERIAL GENES AND THEIR USE

; FILE REFERENCE: 07334/060001

; CURRENT APPLICATION NUMBER: US/09/222.938A

; CURRENT FILING DATE: 1998-12-30

; NUMBER OF SEQ ID NOS: 102

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 15

; LENGTH: 344

; TYPE: DNA

; ORGANISM: Streptococcus pneumoniae

US-09-222-938A-15

Query Match 13.2%; Score 76.6; DB 4; Length 344;

Best Local Similarity 53.7%; Pred. No. 6.4e-14;

Matches 182; Conservative 0; Mismatches 154; Indels 3; Gaps 1;

Qy 113 TGAATCCCGATGTTAGTAAAGGATGATACCCACATTTTCGACCAGCCAGGATATC 172
Db 6 TGTGACCTTGCTGTTGTGGAAGCGTAAATCCATTTCTTCGACCATACCGTTGGAAC 65
Qy 173 GGCCTCTATAAACTTCCCTGTGACCGTTCACGTAACGGAAAGAGTACGTACAACAT 232
Db 66 ACTAATGAAAAGTTTCCCTTGGGATGCTTGGGACGAGGCTAATAAAGATAGATGTC 125
Qy 233 ---AAGTGGGCGAAGAGCTCTCAGAAACGCCAAATCCGGTGACAAAGGGTAGAAGC 289
Db 126 CGCAAGAGGAGAGTAGAAGCTGNAACCCGACATACCGGTATGGAGGATAGAAGC 185
Qy 290 GGTTTAAATTTCCGCCATTTCCGACCGCTGTAGTCTCCGCAAAACGATAAAATGCGTG 349
Db 186 GTCTTAATTTCCACCATTCGACAGCGTGTGTCACGACCTCACTAAAAGCCTAAACGCG 245
Qy 350 GGCACAAATAATAGTCTACCAACCCGATAGAGAGTACAGAAATATGACTGATTTAAAC 409
Db 246 GATAGAAGCAGAGATGGAACGCTAATAGAAGAACCTTGAGAGTATAGAACCCTCATACT 305
Qy 410 AAAGAGAGATAGCTACAAATTTGTCCTAGATATGACAAT 448
Db 306 AAAGTGACAGATCACAGTGTCTAGTAGTACGCCGACAT 344

RESULT 3

US-08-232-463-14/c

; Sequence 14, Application US/08232463

; Patent No. 5670367

; GENERAL INFORMATION:

; APPLICANT: DORNER, F.

; APPLICANT: SCHEIFLINGER, F.

; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232.463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935.313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
; US-08-232-463-14

Query Match. 8.2%; Score 48; DB 1; Length 7218;

Best Local Similarity 2.7%; Pred. No. 8.4e-05;

Matches 9; Conservative 194; Mismatches 129; Indels 0; Gaps 0;

Qy 203 GACGTACGGAAGAGAGTACGTACAACATATAAGTGGGGAAGAGAGCTCTCAGAAACGCC 262
Db 1382 RRR 1323
Qy 263 AAAATCCGGTCCACAAAGGTTAGAGCGGTTTAAATTTCCGCCATTTCCGCCCGCTGTA 322
Db 1322 RRR 1263
Qy 323 GTCCTCCGCAAAACGATAAATGCGTGGGACAAATAAGTGTACCAACGCCATAGA 382
Db 1262 RRR 1203
Qy 383 AGTAGAAAATATGAACGTATTTAAACAAGAGAGAGTAGCTACAATTTGCCCTAGATAT 442
Db 1202 RRR 1143
Qy 443 GACAATATATCAAGAAACAGGTACTATGCTATATAAATACTAAACAGCAATGGGACGAGT 502
Db 1142 RRR 1083
Qy 503 GATAAAACACTATATGCTGTGCTCGCTTG 534
Db 1082 RRR 1051

RESULT 4

US-09-056-075-1/c

; Sequence 1, Application US/09056075
; Patent No. 5955368

GENERAL INFORMATION:

; APPLICANT: Johnson, Eric A.
; APPLICANT: Bradshaw, Marite
; APPLICANT: Rood, Julian
; TITLE OF INVENTION: Expression System for Clostridium
; TITLE OF INVENTION: Species
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53701-2113

COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,075
; FILING DATE:

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

; NAME: Seay, Nicholas J.
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296.95238
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 608-251-5000
; TELEFAX: 608-251-9166

INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 6243 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: 3770..4013

; OTHER INFORMATION: /note= "RP4 origin of DNA transfer (orit) from
; OTHER INFORMATION: plasmid RP4"

US-09-056-075-1

Query Match 6.58; Score 38; DB 2; Length 6243;

Best Local Similarity 49.5%; Pred. No. 0.081;

Matches 98; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

Qy 327 TCCGCAAAACGATAAAATGCGTGGGCAATAAATATAGTGTACCAACGCCATAAGAAGTA 386

Db 1715 TCTAAAGGATATACAGAGAGAGAAATATTAGAGGTAAATAATTTATCAAGAGCAA 1656

Qy 387 GAAAAATATGAATGATTTAAACAAGAGAGAGTAGCTACAAATTTGCCCTAGATATGACA 446

Db 1655 TTAAGAGCTAAAGGAAAGCAACAAAGAGAGGATTAATGTTATTAAGAAAAAATA 1596

Qy 447 ATATATATCAAGAAAGAGGTACTATGCTAATAAATACTACAGCAATGGGACGAGTGATA 506

Db 1595 AAAGCCCTTAAGAAAGAGGCTTTAAATAAATAAGAAATTAATCTAATGTTAGAAGTACCA 1536

Qy 507 AAACACTATATGTCGT 524

Db 1535 ATAAAAACATTAGAACGT 1518

RESULT 5

US-08-353-341-4

; Sequence 4, Application US/08353341

; Patent No. 565417

GENERAL INFORMATION:

; APPLICANT: SUGINAKA, Hidekazu

; APPLICANT: SUGAI, Motoyuki
; APPLICANT: HON, Yonman
; APPLICANT: OGAI, Hideo
; TITLE OF INVENTION: POLYPEPTIDE, PROCESS FOR PREPARING THE
; TITLE OF INVENTION: SAME, AND PHARMACEUTICAL COMPOSITIONS AND COSMETICS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Thomas P. SARRO
; STREET: 727 Twenty-Third Street
; CITY: South Arlington
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22202

COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/353,341
; FILING DATE: 05-DEC-1994
; CLASSIFICATION: 435

PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 689252

; FILING DATE: 12-JUN-1991

; ATTORNEY/AGENT INFORMATION:

; NAME: SARRO, Thomas P.

; REGISTRATION NUMBER: 19396

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 703-920-7200

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 899 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: unknown

; MOLECULE TYPE: DNA (genomic)

US-08-353-341-4

Query Match 6.48; Score 37.2; DB 1; Length 899;

Best Local Similarity 56.6%; Pred. No. 0.066;

Matches 69; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

Qy 372 ACGCATAAGAGTAGAAAAATATGAATGATTTAAACAAGAGAGAGTAGTACAATTG 431

Db 754 ACGCTGTGGAGTGTAGATTGTCAAATGATAAAAGAAAATCATATACACCTATTG 813

Qy 432 TCCCTAGATATGACATATATATATCAAGAAACAGGTACTATGCATAAATACTAACAGCA 491

Db 814 TTTTAAAAATACAAATATAAACAAGATATAACAAGGATTATAAATAATTAATTTTA 873

Qy 492 AT 493

Db 874 AT 875

RESULT 6

US-09-134-001C-1065/c

; Sequence 1055, Application US/09134001C

; Patent No. 6380370

GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCO

; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: GTC-007

; CURRENT APPLICATION NUMBER: US/09/134,001C

; CURRENT FILING DATE: 1998-08-13

; PRIOR APPLICATION NUMBER: US 60/064,964

; PRIOR FILING DATE: 1997-11-08

; PRIOR APPLICATION NUMBER: US 60/055,779

; PRIOR FILING DATE: 1997-08-14

; NUMBER OF SEQ ID NOS: 5674

```

: INFORMATION FOR SEQ ID NO: 1:
:
: SEQUENCE CHARACTERISTICS:
:   LENGTH: 5433 base pairs
:   TYPE: nucleic acid
:   STRANDEDNESS: double
:   TOPOLOGY: linear
:
: MOLECULE TYPE: DNA (genomic)
:
: HYPOTHETICAL: NO
:
: ANTI-SENSE: NO
:
: ORIGINAL SOURCE:
:   ORGANISM: Plasmodium yoelii
:   US-08-929-329-1

```

```

/ APPLICANT: Berglind Ran Olafsdottir
/ APPLICANT: Jeffrey Gulcher
/ TITLE OF INVENTION: HUMAN NARCOLEPSY GENE
/ FILE REFERENCE: 2345.2001-000
/ CURRENT APPLICATION NUMBER: US/09/426,290
/ CURRENT FILING DATE: 1999-10-25
/ NUMBER OF SEQ ID NOS: 24
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 1
/ LENGTH: 168575
/ TYPE: DNA

```


INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1293 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE:
ORIGINAL SOURCE:
FEATURE:
US-08-682-080-1

Query Match 5.6%; Score 32.6; DB 3; Length 1293;
Best Local Similarity 52.6%; Pred. No. 1.8;
Matches 71; Conservative 0; Mismatches 64; Indels 0; Gaps 0;
OY 376 CATAAGAAGTAGAAAATATGAACCTGATTTAAACAAAGAGAGAGTAGCTACAAATGTCCTCC 435
DB 963 CTTAGGACGCTGAAATATGCTGAGAAAACCTGAAAAGGTATAATATTTAGAAATGTCAC 904
OY 436 TAGATATGACAAATATATCAAGAAACAGTACTATGCATATAATCACTAACACGACATGG 495
DB 903 TGTAGGACATGGAAATATAGCAAGAAAACCTGACACTCATGGAATAATGAGAAACATCCCTT 844
OY 496 GACGAGTGATAAAAA 510
DB 843 GACAACTGGAAAAA 829

RESULT 12
US-08-487-826B-13
Sequence 13, Application US/08487826B
Patent No. 5993827
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,826B
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelsen, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CP1
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 19124 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-487-826B-13
Query Match 5.6%; Score 32.6; DB 2; Length 19124;
Best Local Similarity 48.2%; Pred. No. 5.3;
Matches 92; Conservative 0; Mismatches 99; Indels 0; Gaps 0;
OY 392 ATATGAACCTGATTTAAACAAAGAGAGAGTAGCTACAAATGTCCTCCCTAGATATGACAATATA 451
DB 6285 ATATTTGAATATATATATTTATAATGCTATTTTACTTATTTTATTTATATATATACATA 6344
OY 452 TATCAAGAAACAGGTACTATGCATATAATAACTAACGCAATGGGACGAGTGATAAAAAAC 511
DB 6345 AAATTTTGAATTCATAAATGCTGAAATACATAAAAAAATACACAAAAACAATGATA 6404
OY 512 ACTATATGCTGTGGCTCGCTTGTAAATTTGCTTTAATAGTATTTTGTCTTGGATTTCATT 571
DB 6405 AAAACATTTTATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 6464
OY 572 TTACCAATATT 582
DB 6465 TTATCATTTT 6475

RESULT 13
US-08-785-310A-3/C
Sequence 3, Application US/08785310A
Patent No. 5840532
GENERAL INFORMATION:
APPLICANT: McKnight, Steven L.
APPLICANT: Russell, David W.
TITLE OF INVENTION: Neuronal PAS Domain Protein
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP
STREET: 268 BUSH STREET, SUITE 3200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/785,310A
FILING DATE: 21-JAN-1997
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: OSMAN, RICHARD A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: UTSD:1226
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 343-4341
TELEFAX: (415) 343-4342
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 4010 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-785-310A-3

Query Match 5.5%; Score 32.2; DB 2; Length 4010;
Best Local Similarity 57.4%; Pred. No. 3.8;
Matches 56; Conservative 0; Mismatches 43; Indels 0; Gaps 0;
OY 393 TATCAACTGATTTAAACAAAGAGAGAGTAGCTACAAATGTCCTCCCTAGATATGACAATATA 452

[illegible]

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 25, 2002, 11:18:11 ; Search time 171.89 Seconds
(without alignments)
1375.466 Million cell updates/sec

Title: US-10-068-080-12

Perfect score: 582
Sequence: 1 tacaataacgaataacta.....gatttcattttaccaatatt 582

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 363474 seqs, 203117208 residues

Total number of hits satisfying chosen parameters: 726948

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	582	100.0	582	12	US-10-068-080-12
2	100	17.2	642	12	US-10-068-080-11
3	40.4	6.9	739	10	US-09-770-149-62
4	38.6	6.6	4234	10	US-09-070-927A-313
5	37.4	6.4	392	10	US-09-960-352-11703
6	36.8	6.3	1985	9	US-09-905-291A-212
7	36.8	6.3	1985	9	US-09-902-853-212
8	36.8	6.3	1985	10	US-09-909-320-212
9	36.8	6.3	1985	10	US-09-909-088B-212
10	36.8	6.3	1985	12	US-10-052-586-27
11	36	6.2	832	9	US-10-001-857-97
12	36	6.2	27377	10	US-09-816-248-18
13	35.8	6.2	1170	9	US-09-938-842A-5002
14	35.6	6.1	376	10	US-09-960-352-5087
15	35.6	6.1	428	10	US-09-960-352-573
16	35.6	6.1	4167	10	US-09-764-878-282
17	35.6	6.1	4167	10	US-09-764-860-1145
18	35.6	6.1	4167	10	US-09-764-846-345
19	35.4	6.1	357	10	US-09-960-352-1692

ALIGNMENTS

RESULT 1

US-10-068-080-12
Sequence 12, Application US/10068080

Patent No. US20020115591A1

GENERAL INFORMATION:

APPLICANT: Fritz, Christian

APPLICANT: Youngman, Philip

TITLE OF INVENTION: USE OF YNES, ESSENTIAL BACTERIAL GENES AND POLYPEPTIDES

FILE REFERENCE: 07334/090001

CURRENT APPLICATION NUMBER: US/10/068,080

CURRENT FILING DATE: 2002-02-05

PRIOR APPLICATION NUMBER: US 60/070,116

PRIOR FILING DATE: 1997-12-31

NUMBER OF SEQ ID NOS: 12

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 12

TYPE: DNA

ORGANISM: Streptococcus pneumoniae

US-10-068-080-12

Query Match 100.0%; Score 582; DB 12; Length 582;
Best Local Similarity 100.0%; Pred. No. 5.9e-157;
Matches 582; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	TACATTTACGAAATACTAATAAAACCGGATCACTATCCGTCGTAAGCTAGACGCAAT	60
DB	1	TACATTTACGAAATACTAATAAAACCGGATCACTATCCGTCGTAAGCTAGACGCAAT	60
QY	61	TAAACCCGTTGCAAGCGTTTCCTTAACATAAGCCCTCGTCCGCTTGAATCCG	120
DB	61	TAAACCCGTTGCAAGCGTTTCCTTAACATAAGCCCTCGTCCGCTTGAATCCG	120
QY	121	CGATGGTTAGCTAAGCATGTAACCCACATTTTCGACCAAGCAGTAGTATCGGCCTTA	180
DB	121	CGATGGTTAGCTAAGCATGTAACCCACATTTTCGACCAAGCAGTAGTATCGGCCTTA	180
QY	181	TAAACTTTCCCTGTGACCCCTTGACGTAAACGGAAGAGTACGTACAACTATAAGTGGC	240
DB	181	TAAACTTTCCCTGTGACCCCTTGACGTAAACGGAAGAGTACGTACAACTATAAGTGGC	240
QY	241	GAAGAAGCTCCTCAGAAACCCAAAATCCGGTGCACAAAGGTAGTAAGACGGCTTTAAATTT	300

C	20	35.4	6.1	446	10	US-09-960-352-3400	Sequence 3400, Ap
C	21	35.4	6.1	516	10	US-09-960-352-5785	Sequence 5785, Ap
C	22	35.4	6.1	4237	10	US-09-745-763-20	Sequence 20, Appl
C	23	35.4	6.1	21636	10	US-09-416-384A-3	Sequence 3, Appl
C	24	35	6.0	414	10	US-09-960-352-6528	Sequence 6528, Ap
C	25	35	6.0	442	10	US-09-960-352-12911	Sequence 12911, A
C	26	35	6.0	480	10	US-09-960-352-5301	Sequence 5301, Ap
C	27	34.8	6.0	426	10	US-09-960-352-8406	Sequence 8406, Ap
C	28	34.8	6.0	463	10	US-09-960-352-7186	Sequence 7186, Ap
C	29	34.8	6.0	21860	10	US-09-764-877-3803	Sequence 3803, Ap
C	30	34.8	6.0	65359	10	US-09-804-472-3	Sequence 3, Appl
C	31	34.6	5.9	385	10	US-09-960-352-1739	Sequence 1739, Ap
C	32	34.6	5.9	431	10	US-09-960-352-5558	Sequence 5558, Ap
C	33	34.4	5.9	377	10	US-09-960-352-5583	Sequence 5583, Ap
C	34	34.4	5.9	1198	10	US-09-789-561-55	Sequence 55, Appl
C	35	34.2	5.9	484	10	US-09-817-607-10	Sequence 10, Appl
C	36	34.2	5.9	1959	10	US-09-070-927A-282	Sequence 282, App
C	37	34	5.8	352	10	US-09-960-352-10129	Sequence 10129, A
C	38	34	5.8	1200	10	US-09-887-576-758	Sequence 758, App
C	39	34	5.8	2000	9	US-09-938-842A-3681	Sequence 3681, Ap
C	40	33.8	5.8	415	10	US-09-960-352-2223	Sequence 2223, Ap
C	41	33.8	5.8	448	10	US-09-960-352-11904	Sequence 11904, A
C	42	33.8	5.8	451	10	US-09-960-352-10262	Sequence 10262, A
C	43	33.8	5.8	948	9	US-09-938-842A-2359	Sequence 2359, Ap
C	44	33.6	5.8	259	10	US-09-960-352-7233	Sequence 7233, Ap
C	45	33.6	5.8	439	10	US-09-960-352-1009	Sequence 1009, Ap

```

Db 241 GAAGAACGTCCTCAGAAACGCCAAAATCCGGTCACAAAGGGTAGAAGCGGTTTAAATTT 300
QY 301 CCGCCATTTCCGACCCGCTGACTCTCCGCAAAAGCATAAATCCGTCGGGACAAATATA 360
Db 301 CCGCCATTTCCGACCCGCTGACTCTCCGCAAAAGCATAAATCCGTCGGGACAAATATA 360
QY 361 TAGTGTCTACCAACGCCATAAGAGTAGAAAAATATGAACCTGATTTAAACAAAGAGAGAT 420
Db 361 TAGTGTCTACCAACGCCATAAGAGTAGAAAAATATGAACCTGATTTAAACAAAGAGAGAT 420
QY 421 AGCTACAATTTGCTCCCTAGATATGACAAATATATATCAAGAAACAGAGTACTATGATAAAT 480
Db 421 AGCTACAATTTGCTCCCTAGATATGACAAATATATATCAAGAAACAGAGTACTATGATAAAT 480
QY 481 AACTAACAGCAATGGGACGAGTATATAAAACACTATATGCTGTGCTCCCTGTTAATTT 540
Db 481 AACTAACAGCAATGGGACGAGTATATAAAACACTATATGCTGTGCTCCCTGTTAATTT 540
QY 541 GCTTAATAGTATTGCTGCTTGGATTTCATTTTACCATAATT 582
Db 541 GCTTAATAGTATTGCTGCTTGGATTTCATTTTACCATAATT 582

RESULT 2
US-10-068-080-11
; Sequence 11, Application US/10068080
; Patent No. US20020115591A1
; GENERAL INFORMATION:
; APPLICANT: Fritz, Christian
; APPLICANT: Youngman, Philip
; TITLE OF INVENTION: USE OF YNES, ESSENTIAL BACTERIAL GENES AND POLYPEPTIDES
; FILE REFERENCE: 07334/090001
; CURRENT APPLICATION NUMBER: US/10/068,080
; CURRENT FILING DATE: 2002-02-05
; PRIOR APPLICATION NUMBER: US 60/070,116
; PRIOR FILING DATE: 1997-12-31
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 11
; LENGTH: 642
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-10-068-080-11

Query Match 17.2%; Score 100; DB 12; Length 642;
Best Local Similarity 52.9%; Pred. No. 3.1e-19;
Matches 239; Conservative 0; Mismatches 210; Indels 3; Gaps 1;

QY 1 TACAATTAACGAATAACTAATAAAACCGGATGAACCTATCCGTCGTAAGGTAGACCGAAT 60
Db 1 TACTAATGTTATCAAAATATATTAGGATCGGATAGACGACCCCAAGCTAAGGTAGACCAAG 60
QY 61 TAACACCGGTCGAACGGTTTCCTTAACATAAGCCCTCGTGCTCGCCGTTGGAATCCG 120
Db 61 ACCTAACCTGTTCTATAAGAAAGTTTAGTTAGATCGCTCGTACCAAGACCAATTTGACCT 120
QY 121 CGATGGTTAGTAAAGCATGTACCCACATTTTCGACCAAGCAGCATGATCGCCCTCTA 180
Db 121 TCGTGGTTGTGGAAGGCGTAAATCTCTTTTCGACCATACCTGTTGGAAACACTAACTG 180
QY 181 TAAACCTTTCTCTGACCGCTTACGTTAAGCGAAAGAGTACGTACA- --ACTATAAGTG 237
Db 181 AAAAGTTTCTCTGGGATCGTTGCGACGAGGCTAATAAAGTAGATGTTCCCAAGA 240
QY 238 GCGAAGAACGCTCTCAGAACGCCAAATCCGGTCGACAAAGGTTAGAACGGTTTAA 297
Db 241 GGAGAGTAGAAACCTGAAAACCGCAATAGCCGGTATGGAAGGGATAGAAACGCTCTTAA 300
QY 298 TTTCCGCAATTTCCGACCGCTAGTCTCCCGCAAAACGATATAAATCGCTGGGCAAT 357
Db 301 TTTCCACCATTTCCGACACGCGTTGCTCACGACCTCACCTAAAGCCCTAAACCGGATAGA 360
```

```

QY 358 AAATAGTGTCTACCAACGCCCATAGAAGTAGAAAAATATGAACCTGATTTTAAACAAAGAGAG 417
Db 361 ACAGAGATGGAACGCTAATAAGAGAAACCTCGAGAGATAGAACCGGTCTACTATAAGCTGAC 420
QY 418 AGTAGCTACAAATTTGTCCTCTAGATATGACAATA 449
Db 421 AGATCACAGTGTGCTAGCTTAACGCCGACAATA 452

RESULT 3
US-09-770-149-62
; Sequence 62, Application US/09770149
; Patent No. US20020059663A1
; GENERAL INFORMATION:
; APPLICANT: Gorlach, Jorn
; APPLICANT: An, Yong-Qiang
; APPLICANT: Hamilton, Carol M.
; APPLICANT: Price, Jennifer L.
; APPLICANT: Raines, Tracy M.
; APPLICANT: Yu, Yang
; APPLICANT: Rameaka, Joshua G.
; APPLICANT: Page, Amy
; APPLICANT: Matthew, Abraham V.
; APPLICANT: Ledford, Brooke L.
; APPLICANT: Woessner, Jeffrey P.
; APPLICANT: Haas, William David
; APPLICANT: Garcia, Carlos A.
; APPLICANT: Krieker, Maja
; APPLICANT: Slader, Ted
; APPLICANT: Davis, Keith R.
; APPLICANT: Allen, Keith
; APPLICANT: Hoffman, Neil
; APPLICANT: Hurlban, Patrick
; TITLE OF INVENTION: Expressed Sequences of Arabidopsis
; TITLE OF INVENTION: thaliana
; FILE REFERENCE: 2024 (PARA-013PRV)
; CURRENT APPLICATION NUMBER: US/09/770,149
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 60/178,506
; PRIOR FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 999
; SOFTWARE: FastSeq for Windows, Version 4.0
; SEQ ID NO 62
; LENGTH: 739
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(739)
; OTHER INFORMATION: n = A,T,C or G
US-09-770-149-62

Query Match 6.9%; Score 40.4; DB 10; Length 739;
Best Local Similarity 53.9%; Pred. No. 0.035;
Matches 83; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 364 TGCTACCAACGCCCATAGAAGTAGAAAAATATCAACCTGATTTTAAACAAAGAGAGTAGC 423
Db 286 TGCATATAATTCAGCAAACTTCGGTAAAGATAATTTGTAATTTGAAGAGTCCATAAC 345
QY 424 TACAATTTGTCCTCTAGATATGACAATATATATCAAGAAACAGAGTACTATGCAATAAATAC 483
Db 346 TGATGGTCTACCATATTGAAACTCTAAACCATAGAACCCAGCGCAAAAGCAATTAA 405
QY 484 TACACCAATGGGACGAGTGTATATAAACACTATA 517
Db 406 CATCAGCATAGGAAAAAGAGCTATAATACCAAA 439

RESULT 4
US-09-070-927A-313
; Sequence 313, Application US/09070927A
; Patent No. US20020120116A1
```

```

; GENERAL INFORMATION:
; APPLICANT: Charles A. Kunsch
; Patrick J. Dillion
; Steven Barash
; TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 982
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/070,927A
; FILING DATE: 04-May-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/046,655
; FILING DATE: 1997-05-16
; APPLICATION NUMBER: 60/044,031
; FILING DATE: 1997-05-06
; APPLICATION NUMBER: 60/066,009
; FILING DATE: 1997-11-14
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenley K. Hoover
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB369
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 313:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4234 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 313:
US-09-070-927A-313

Query Match 6.6%; Score 38.6; DB 10; Length 4234;
Best Local Similarity 50.8%; Pred. No. 0.25; Mismatches 89; Indels 0; Gaps 0;
Matches 92; Conservative 0;

QY 330 GCAAAACGATAAAATGCGTGGGACAAATAAATAGTGTCTACCAACGCCATAAGAGTAGAA 389
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1886 GCAAACTTAAAGTAAAGCGGCAGACCCCAATGTGACCATTTGGAAGTAACATAATC 1945
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 390 AATATGAAGTATTTAAACAAGAGAGAGTAGCTACAATTGTCCTAGATATGACAATA 449
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1946 AATATGTAAGAAAGTAACAAACAAGTAATAGACAAAATAAAATCCCACTGGCAGTA 2005
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 450 TATATCAAGAACAGGTACTATGTCATTAATAACTAACGCAATGGAGCAGTGATAAA 509
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2006 CCAATTACAGATTATAGGCTTTTAAAGGCCATACCTTAAGGCATTGGGCAAGTAATATTA 2065
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 510 A 510
Db 2066 A 2066

RESULT 5
US-09-960-352-11703/c
; Sequence 11703, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
```

```

; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION A
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 11703
; LENGTH: 392
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 50-LIB3058-031-01-X1-E6
US-09-960-352-11703

Query Match 6.4%; Score 37.4; DB 10; Length 392;
Best Local Similarity 57.8%; Pred. No. 0.19; Mismatches 61; Indels 1; Gaps 1;
Matches 85; Conservative 0;

QY 371 AACGCGCATAGAGTAGAAAAATATCACTGATTTAAACAAGAGAGAGAGTAGCTACAATT 430
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 368 AATTACAAAAAGAAAAAATAAATACTAAAAAATAAGACA-AAAAAAAAAATG 310
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 431 GTCCTAGATATGACAATATATATCAAGAAACAGGTACTATGCATAATACTAACAGC 490
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 309 AATAAAAAATATAAAAAATATAGAAAAATAAAGAAAAATAAAAAACAATAAAAAA 250
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 491 AATGGCAGGTGATAAAAAACACTATA 517
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 249 AAAAAATAAAAAATAAAAAAATAAAAAATAAA 223
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 6
US-09-905-291A-212
; Sequence 212, Application US/09905291A
; Patent No. US20020160374A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/905,291A
; CURRENT FILING DATE: 2001-07-12
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
```

RESULT 8

```
US-09-909-320-212
; Sequence 212, Application US/09909320
; Patent No. US20020132240A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/909,320
; PRIOR FILING DATE: 2002-01-04
; CURRENT APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: PCT/US99/28313
; PRIOR FILING DATE: 1999-11-30
; PRIOR APPLICATION NUMBER: PCT/US99/28564
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/28565
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30095
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: PCT/US99/30911
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US99/30999
; PRIOR FILING DATE: 1999-12-20
; PRIOR APPLICATION NUMBER: PCT/US00/00219
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 423
; SEQ ID NO 212
; LENGTH: 1985
; TYPE: DNA
; ORGANISM: Homo sapiens

US-09-909-320-212
Query Match 6.3%; Score 36.8; DB 10; Length 1985;
Best Local Similarity 51.9%; Pred. No. 0.58; Mismatches, 77; Indels 0; Gaps 0;
Matches 83; Conservative 0;

QY 351 GGACAATAAATAGTGTACCAACGCCATAGAGAGTAAAGAAATATGAACTGATTTAAACA 410
Db 1826 GGACAATAAATAGTGTATGACTGCTTCGCCCAAAAAAAAAAAAAAAAAAAAAA 1885

QY 411 AAGAGAGAGTACGTACAATTGTCCTAGATATGACATATATATCAAGAAAGACAGGTACT 470
Db 1886 AAAAAAATAAATAGTGTATGACTGCTTCGCCCAAAAAAAAAAAAAAAAAAAAAA 1945

QY 471 ATGCATAAATAACTAAACAGCAATGGGACGAGTGATATAAAAA 510
Db 1946 AAAAAAATAAATAGTGTATGACTGCTTCGCCCAAAAAAAAAAAAAAAAAAAAAA 1985

RESULT 9
US-09-909-088B-212
; Sequence 212, Application US/09909088B
; Patent No. US20020146709A1
; GENERAL INFORMATION:
; APPLICANT: Genentech, Inc.
; APPLICANT: Ashkenazi, Avi
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, A.
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Hillan, Kenneth, J.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Mather, Jennie P.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William, I.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: 10466-14
; CURRENT APPLICATION NUMBER: US/09/909,088B
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: PCT/US00/04414
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 60/143,048
; PRIOR FILING DATE: 1999-07-07
; PRIOR APPLICATION NUMBER: US 60/145,698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: US 60/146,222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: PCT/US99/20594
; PRIOR FILING DATE: 1999-09-08
; PRIOR APPLICATION NUMBER: PCT/US99/20944
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: PCT/US99/21090
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/21547
; PRIOR FILING DATE: 1999-09-15
; PRIOR APPLICATION NUMBER: PCT/US99/23089
; PRIOR FILING DATE: 1999-10-05
; PRIOR APPLICATION NUMBER: PCT/US99/28214
```

APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC ACIDS ENCODING THE SAME

```

; PRIOR FILING DATE: 1998-06-10
; PRIOR APPLICATION NUMBER: 60/088861
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088863
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/088876
; PRIOR FILING DATE: 1998-06-11
; PRIOR APPLICATION NUMBER: 60/089090
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089105
; PRIOR FILING DATE: 1998-06-12
; PRIOR APPLICATION NUMBER: 60/089512
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089514
; PRIOR FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: 60/089538
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089598
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089653
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089908

Query Match          6.3%; Score 36.8; DB 12; Length 1985;
Best Local Similarity 51.9%; Pred. No. 0.58;
Matches 83; Conservative 0; Mismatches 77; Indels 0; Gaps

QY 351 GCACAATAAATAGTCTACCAACGCCATAAGCAAGTAGAGAAAATATGAACTGATTTAAACA 410
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Ddb 1825 GGCACATAATAGTGCTATGACTGCCTCCGCCAAAANAANAANAANAANAANAANA 1885
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 411 AAAGAGAGAGTAGCTACAATTGCTCCCTAGATATGACAATATATATCAAGAAACAGGTACT 470
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Ddb 1886 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1945

QY 471 ATGCATAATAACTAACAGCAATGGGACGAGTGATAAAAA 510
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Ddb 1946 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1985

RESULT 11
US-10-001-857-97
; Sequence 97, Application US/10001857
; Publication No. US20020183500A1
; GENERAL INFORMATION:
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Herve
; APPLICANT: Chen, Sei-Yu
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific G
; FILE REFERENCE: DEX-0273
; CURRENT APPLICATION NUMBER: US/10/001.857
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/252,054
; PRIOR FILING DATE: 2000-11-20
; NUMBER OF SEQ ID NOS: 208
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 97
; LENGTH: 832
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-001-857-97

Query Match          6.2%; Score 36; DB 9; Length 832;
Best Local Similarity 49.5%; Pred. No. 0.67;
Matches 93; Conservative 0; Mismatches 95; Indels 0; Gaps

QY 330 GCACAACGATAAATCCGTGGGCAATAAATAGTGTACCACGCCCATAGAAGTAGAA 389
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Ddb 628 GGACAAGAACCAAGAGATAAGGAAAAAACAACAAAGACGACGAAGGAAAAATAGAGAGA 687
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 390 AAATATGAAGTATTAAACAAGAGAGAGTAGCTACAAATTTGTCCTAGATATGACAATA 449

```

```

Db 688 AACCAACAAACAAACAAACAAACAAACCAACGACGAGAAACAACTAGAGAAA 747
Qy 450 TATATCAAGAACAGCTACTATGTCATAATAATACTAACACCAATGGGACGAGTATAAAA 509
Db 748 ACAAAAGAACACAAACAAACAAACCAACCAACCAACCAACCAACCAACCAACCAAC 807
Qy 510 ACACATATA 517
Db 808 GCACGAAA 815

RESULT 12
US-09-816-248-18/c
; Sequence 18, Application US/09816248
; Patent No. US20020137703A1
; GENERAL INFORMATION:
; APPLICANT: BAUMANN, PETER
; APPLICANT: CECH, THOMAS R.
; TITLE OF INVENTION: PROTECTION-OF-TELOMERE-1 (POT-1) PROTEIN AND ENCODING
; FILE REFERENCE: 089491/0201
; CURRENT APPLICATION NUMBER: US/09/816,248
; CURRENT FILING DATE: 2001-03-26
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 27377
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-816-248-18

Query Match 6.2%; Score 36; DB 10; Length 27377;
Best Local Similarity 51.2%; Pred. No. 3.2;
Matches 84; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

Qy 414 AGACAGTAGCTACAAATGTCCTAGATATGCAATATATATCAAAAGAACAGGTACTATG 473
Db 1179 AGCTCGGAGGACGAGGAGACTCTCAAAAGAAAAAAGAGAGATATG 1120
Qy 474 CATAAATACTACAGCAATGGGACGAGTGATAAAACATATATGTCCTGGCTGCCTT 533
Db 1119 TAGAAATAAATACAGTCGACAGACACACAAAGTAAATAGTTATTGGAGAGTCTAGCTG 1060
Qy 534 GTAATTTGCTTAATAGTATTTTCTCTGATTTTACCA 577
Db 1059 GTAGTTATTGAGTATTCATTTGACAACTGTTTAAATTTACCA 1016

RESULT 13
US-09-938-842A-5002/c
; Sequence 5002, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kieps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; FILE REFERENCE: SCRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 5002
; LENGTH: 1170
; TYPE: DNA
```

```

; ORGANISM: Arabidopsis thaliana
US-09-938-842A-5002

Query Match 6.2%; Score 35.8; DB 9; Length 1170;
Best Local Similarity 55.1%; Pred. No. 0.89;
Matches 70; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

Qy 364 TGCTACCAACGCCATAGAAGTAGAAAAATATCAACTGATTTTAAACAAGAGAGAGTAGC 423
Db 159 TGTACGATCATATATATAGTAAATAAAACATATCATTTTACAAAAAAGAGAGG 100
Qy 424 TACAATTGTCCTAGATATGACAAATATATCAAGAAGACAGGTACTATGCATAAATAAC 483
Db 99 AAAAATAAAACGCAAAACCCCTAGTGTATTTGATACAAAACCCCTAGTATATATAAAGAAC 40
Qy 484 TAACAGC 490
Db 39 AGAAAAC 33

RESULT 14
US-09-960-352-5087/c
; Sequence 5087, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION A
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 5087
; LENGTH: 376
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 22-LIB3058-032-Q1-K1-F9
US-09-960-352-5087

Query Match 6.1%; Score 35.6; DB 10; Length 376;
Best Local Similarity 49.5%; Pred. No. 0.61;
Matches 92; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

Qy 332 AAAACGATAAAATGCGTGGGACAAATAATAGTGTACCAACCCCATAGAGAGAGAGAA 391
Db 348 AAAAATAAAAAATATAACAAAATAACAAATAATAAAAAAATAAAAAAATAAAAA 289
Qy 392 ATATGAACTGATTTAAACAAGAGAGAGTAGCTACAAATTTGTCCTCCTAGATATGCAATATA 451
Db 288 TAATAAATAATAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAA 229
Qy 452 TATCAAGAAACAGGTACTATGCAATAAATACTAACAGCAATGGGACGAGTATAAAAAAC 511
Db 228 AATTAAAAATAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATA 169
Qy 512 ACTATA 517
Db 168 AAAATA 163

RESULT 15
US-09-960-352-573/c
; Sequence 573, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION A
```

```

; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 573
; LENGTH: 428
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 03-LIB3058-031-Q1-K1-A11
US-09-960-352-573

```

Query Match 6.1%; Score 35.6; DB 10; Length 428;

Best Local Similarity 49.5%; Pred. No. 0.64;

Matches 92; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

```

Qy 332 AAACGATAAATCCGTGGGACAAATTAATAGTCTACCAACGCCATAAGAAGTAGAAAA 391
    |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||
Db 376 AAAAAAAAAAAAAAAAAAAAAAAAAAATATAAAAAAAAAAAAAAAAAAGAAATATA 317
    |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||
Qy 392 ATATGAACCTGATTAAACAAAGAGAGAGTAGCTACAAATTCCTCTAGATATGACAATATA 451
    |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||
Db 316 AAAAAAAAAAGAAATATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATA 257
    |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||
Qy 452 TATCAAGAAACAGGTACTATGCATAAATACTAACAGCAATGGGAGCTGATATAAAC 511
    |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||
Db 256 AAAAAAAAAAAAAAAAAAAAAAAAAAATATAAAAAAAAAAAAAAAAAAGAAATATA 197
    |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||
Qy 512 ACTATA 517
    |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||
Db 196 AAGAAA 191
    |||| | |||| | |||| | |||| | |||| | |||| | |||| | |||| | ||||

```

Search completed: December 26, 2002, 00:46:35
Job time : 201.89 secs